П. БИЛЛИНГСЛЕЙ

ЭРГОДИЧЕСКАЯ ТЕОРИЯ И ИНФОРМАЦИЯ

ИЗДАТЕЛЬСТВО
«МИР»

Ergodic Theory and Information

PATRICK BILLINGSLEY
The University of Chicago

John Wiley and Sons, Inc.
New York • London • Sydney
1965

П. БИЛЛИНГСЛЕЙ

Эргодическая теория и информация

Перевод с английского
Н. Д. СВЕТЛОВОЙ

Под редакцией
Б. М. ГУРЕВИЧА

ИЗДАТЕЛЬСТВО „МИР“ МОСКВА 1969

УДК 519.21

Эта книга посвящена в основном новейшим результатам эргодической теории, связанным в первую очередь с теоретико-информационными методами исследования динамических систем. Изложив основные понятия эргодической теории (сохраняющего меру преобразования, эргодичности, перемешивания и т. п.), автор переходит к определению энтропии динамической системы. Далее вводится понятие условной энтропии, описываются ее свойства, дается теорема Макмиллана и рассматриваются связи между понятием размерности в смысле Хаусдорфа и понятием энтропии. В заключение излагаются основные положения общей теории связи в смысле Шеннона и обсуждаются соответствующие вопросы эргодической теории.

Книга написана ясно и не требует от читателя большой математической подготовки. Она, безусловно, заинтересует математиков многих специальностей.

Редакция литературы по математическим наукам

Инд. 2-2-3
4-68

Предисловие редактора перевода

Теория, основам которой посвящена эта книга, отметила в 1968 г. свое десятилетие. Введение А. Н. Колмогоровым десять лет назад понятия энтропии преобразования с инвариантной мерой не только позволило ему решить давно поставленную конкретную задачу, но положило начало новому — энтропийному—направлению эргодической теории, ставшему вскоре в ней ведущим. Даже последующие достижения в традиционных для эргодической теории вопросах обязаны энтропийному направлению возрождением интереса математиков к эргодической теории в целом.

Книга профессора Биллингслея, написанная с большим педагогическим мастерством, представляет собой первое на русском языке изложение основ энтропийной теории преобразований с инвариантной мерой, доступное начинающему. От читателя требуется лишь владение абстрактным интегралом Лебега и элементарными понятиями теории вероятностей. Автор подробно излагает теорию условных вероятностей, и соответствующие параграфы можно рекомендовать для первоначального ознакомления с предметом даже тем из читателей, кого основное содержание книги не заинтересует.

Хотя перед нами книга учебного характера и в нее не включены многие глубокие результаты энтропийной теории (с этими результатами можно познакомиться по обзорной статье В. А. Рохлина „Лекции по энтропийной теории преобразований с инвариантной мерой“, *УМН*, **5** (1967)), кое-что в ней может заинтересовать и специалиста, в частности помещенная в гл. 3 теорема Биллингслея о связи энтропии сдвига с хаусдорфовой размерностью.

В книге разобрано довольно много примеров. Одному из них — эндоморфизму, возникающему в связи с представлением вещественного числа непрерывной дробью, посвящен заключительный параграф гл. 1. Здесь имеется компактное изложение основных результатов метрической теории непрерывных дробей, написанное с „эргодической“ точки

зрения. Почти все остальные примеры вероятностного происхождения, и поэтому у читателя может возникнуть впечатление, что энтропийная теория имеет дело преимущественно лишь с такого рода примерами. В действительности же не менее важной и интересной областью ее применения являются диффеоморфизмы гладких многообразий.

В небольшом приложении, написанном Я. Г. Синаем и автором этих строк и помещенном в конце книги, рассматривается один класс таких преобразований—алгебраические автоморфизмы n-мерного тора. Материал этого приложения поможет читателю получить представление о характере применения энтропии и связанных с ней понятий к гладким преобразованиям. Используемый метод основан на понятии марковского разбиения, недавно введенном и изученном Я. Г. Синаем.

Можно надеяться, что выход в свет книги Биллингслея расширит круг лиц, интересующихся эргодической теорией, и окажет помощь всем, кто захочет ее изучить.

Б. Гуревич

Предисловие к английскому изданию

Мне очень приятна возможность в нескольких словах представить книгу Патрика Биллингслея, первую в новой серии «Монографий по теории вероятностей и математической статистике» (Tracts on Probability and Mathematical Statistics).

Эта книга возникла из цикла лекций, прочитанных профессором Биллингслеем на учебных конференциях Лондонского математического общества, лекций, которые, как нам кажется, должны стать достоянием более широкой публики. Особенно удачно, что новую серию открывает именно эта книга, поскольку и «Монографии» и конференции Лондонского математического общества стремятся сделать новые математические достижения общедоступными еще на той стадии, когда в них бьется живой пульс, прежде чем они примут застывшую форму.

Мы надеемся сохранить умеренный объем всех книг этой серии. Большинство возможных авторов отпугивает мысль о необходимости написать, отпечатать и прокорректировать сотни страниц математического текста. Нет сомнения в том, что много прекрасных математических книг никогда не были написаны именно по этой причине, а также потому, что для тех, кто должны были быть их авторами, оказалось невозможным оторваться от преподавания и исследовательской работы на срок, необходимый для написания исчерпывающего трактата.

Мы надеемся также внести этой серией вклад в улучшение существующих стандартов в области планирования и издания математических книг. Книга бесполезна, если ее не читают. Издатели убеждены, что недостаточно только захватывающего сюжета и мастерства; сам текст должен быть представлен привлекательно, что определенно способствует пониманию рассуждений. Это вопрос не эстетики, а доступности, и мы думаем, что он заслуживает большего внимания.

Кембридж, Англия

Дэвид Кендалл

Введение

Эта книга выросла из обзора[1]) эргодической теории и теории информации, который должен был следовать за кратким курсом теории вероятностей на основе теории меры.

В гл. 1 рассматривается эргодическая теорема. Глава 2 посвящена данному Колмогоровым и Синаем приложению шенноновской энтропии к проблеме изоморфизма в эргодической теории. В противоречии с исторической последовательностью эти идеи только потом (в гл. 4 и 5) применяются к теории информации и теории кодирования. Результаты, относящиеся к кодированию, не претендуют на особую глубину; я ставил своею целью возможно теснее связать теорию кодирования и эргодическую теорию, избегая технических подробностей. Связь между этими двумя теориями является той нитью, которая связывает все параграфы книги.

С самого начала предполагается, что читатель знаком с пространством $(\Omega, \mathcal{F}, P)$. Впрочем, чтобы сократить ссылки на теорию меры, я включил (гл. 3) описание условных вероятностей и математических ожиданий относительно σ-поля и отметил звездочкой те темы, которые могут быть опущены (конечно, звездочкой могла бы быть отмечена вся книга).

Я пытался писать не для знатоков, а для новичков. Я пытался следовать блестящему примеру Харди и Райта, написавших свое «An introduction to the theory of numbers» с заранее поставленной целью (указанной в их предисловии) создать интересную книгу.

Копенгаген, декабрь 1964 г. *Патрик Биллингслей*

---

[1]) Курс лекций, представленных Лондонскому математическому обществу на учебной конференции по теории вероятностей, проходившей в Дёрем-Колледж с 28 марта по 11 апреля 1963 г.

ГЛАВА 1

Эргодическая теория

1. ПРЕОБРАЗОВАНИЯ, СОХРАНЯЮЩИЕ МЕРУ

Введение

Случай необходимо связан с понятием изменчивости, но сами управляющие изменчивостью законы могут оставаться неизменными во времени: если рулетка не стареет со временем, счастье улыбается игроку в соответствии с постоянными вероятностными законами. Эргодическая теория дает ключ к пониманию этих случайных изменений.

Бросание монеты, игральной кости, наблюдение длины очереди, определение числа молекул в данном объеме — вообразите любой из таких случайных экспериментов или наблюдений. Пространство состояний эксперимента есть множество ρ его возможных исходов: ρ состоит из герба и решетки, граней кости и т. п. Пусть наш эксперимент производится, скажем, раз в минуту, и это продолжалось и будет продолжаться вечно.

Бесконечную в обе стороны последовательность экспериментов можно рассматривать как один большой эксперимент, исход которого представляет собой бесконечную в обе стороны последовательность $\omega = (\ldots, \omega_{-1}, \omega_0, \omega_1, \ldots)$ элементов множества ρ. Вероятностная структура этого большого эксперимента задается вероятностной мерой P на пространстве Ω таких последовательностей[1]).

Допуская возможность того, что исход каждого эксперимента сильно влияет на исходы следующих за ним, мы хотим математически выразить идею, что течение времени не влияет на совместные распределения вероятностей, которым подчиняется экспериментирование. Сдвигая последовательность ω влево на один шаг, получаем новую последовательность $\omega' = (\ldots, \omega_0, \omega_1, \omega_2, \ldots)$, в которой ω_1 стоит на нулевом месте. Так как ω и ω' — идентичные реализации

---

[1]) Трудно представить, как согласовать P с частотной интерпретацией теории вероятностей: большой эксперимент не может быть повторен, так как проведение его занимает все время. Однако хотя бесконечность последовательности компонент эксперимента математически существенна, в реальных ситуациях она не должна пониматься буквально.

большого эксперимента, отличающиеся только началом отсчета времени, то вероятностная мера P должна приписывать ω' ту же вероятность, что и ω, если вероятностные законы неизменны во времени. На самом деле ω и ω' будут иметь, вообще говоря, вероятность 0, и нужно требовать, чтобы при преобразовании T, переводящем ω в ω', вероятностная мера P сохранялась в том смысле, что $P(A)=P(TA)$ для множеств A. Это приводит нас к изучению преобразований, сохраняющих меру, — к эргодической теории.

В этом параграфе, после того как будут даны первые определения, примеры (как из теории вероятностей, так и из других областей математики) и общие принципы, мы сформулируем эргодическую теорему, представляющую собой основной результат теории, и приведем примеры ее применения. В § 2 содержатся доказательства эргодической теоремы, а в § 3 и 4 — дальнейшие примеры и применения.

Определения

В эргодической теории изучаются преобразования, сохраняющие структуру пространств с мерой. Мы будем заниматься эргодической теорией в ее связи с теорией вероятностей и теорией информации и потому ограничимся рассмотрением вероятностных пространств.

Пусть $(\Omega, \mathscr{F}, P)$ — вероятностное пространство[1]) и T — преобразование, отображающее Ω в себя, *измеримое* в том смысле, что из $A \in \mathscr{F}$ следует $T^{-1}A = \{\omega: T\omega \in A\} \in \mathscr{F}$. Если, кроме того, T — взаимно однозначное преобразование, $T\Omega = \Omega$ и из $A \in \mathscr{F}$ следует $TA = \{T\omega: \omega \in A\} \in \mathscr{F}$, то T называется *обратимым*. В случае когда $P(T^{-1}A) = P(A)$ для каждого A из $\mathscr{F}$, говорят, что T — *сохраняющее меру* преобразование; если T обратимо, то эквивалентным требованием является $P(TA) = P(A)$[2]).

(Даже если сохраняющее меру преобразование необратимо, $T\Omega$ есть по существу все Ω, так как $T\Omega \subset A \in \mathscr{F}$ влечет за собой $T^{-1}A = \Omega$ и, следовательно, $P(A) = 1$. В частности, если $T\Omega$ принадлежит $\mathscr{F}$, то $P(T\Omega) = 1$.)

---

[1]) Ω — пространство точек ω, $\mathscr{F}$ — это σ-поле (т. е. σ-алгебра) подмножеств из Ω, P — вероятностная мера на $\mathscr{F}$. Об аксиоматическом построении теории вероятностей см., например, Колмогоров [1] или Халмош [2, гл. X].

[2]) В отечественной литературе принято называть измеримое сохраняющее меру преобразование *эндоморфизмом*, а если оно к тому же обратимо, *автоморфизмом*. — *Прим. ред.*

Примеры

Рассмотрим сначала два вероятностных примера.

Пример 1.1.[1]) Пусть ρ — конечное множество из r элементов. Приписывая элементам множества ρ неотрицательные вероятности p_i так, что $\sum_{i\in\rho} p_i = 1$, мы получаем вероятностную меру на ρ (вернее на σ-поле всех подмножеств множества ρ). Пусть $(\Omega, \mathscr{F}, P)$ есть произведение $\prod_{n=-\infty}^{\infty} \rho_n$ пространств ρ_n, каждое из которых совпадает с ρ. Произвольный элемент пространства Ω является бесконечной в обе стороны последовательностью $\omega = (\ldots, \omega_{-1}, \omega_0, \omega_1, \ldots)$ элементов множества ρ. Пусть x_n — функция, ставящая в соответствие точке $\omega \in \Omega$ значение ω_n её n-й координаты. Функцию x_n называют n-й координатной функцией. Конечно аддитивное поле, состоящее из цилиндров, т. е. множеств вида

$$\{\omega\colon (x_n(\omega), \ldots, x_{n+k-1}(\omega)) \in E\} = \{\omega\colon (\omega_n, \ldots, \omega_{n+k-1}) \in E\},$$

где E — подмножество прямого произведения ρ^k, порождает σ-поле $\mathscr{F}$. Но σ-поле $\mathscr{F}$ порождается также совокупностью множеств, которые мы назовем „тонкими“ цилиндрами, т. е. множеств вида $\{\omega\colon x_l(\omega) = i_l,\ n \leqslant l < n+k\}$, где i_l — элементы множества ρ; в самом деле, каждый цилиндр является конечным объединением непересекающихся тонких цилиндров. Наконец, мера P определяется совокупностью ее значений на тонких цилиндрах[2]):

$$P\{\omega\colon x_l(\omega) = i_l,\ n \leqslant l < n+k\} = \prod_{l=n}^{n+k-1} p_{i_l}. \tag{1.1}$$

Итак, $\{\ldots, x_{-1}, x_0, x_1, \ldots\}$ является последовательностью независимых случайных величин со значениями из ρ. Пусть $T\colon \Omega \to \Omega$ — отображение, переводящее $(\ldots, \omega_{-1}, \omega_0, \omega_1, \ldots)$ в $(\ldots, \omega_0, \omega_1, \omega_2, \ldots)$. Более точно T определяется соотношением $(T\omega)_n = \omega_{n+1}$, или (что то же самое) $x_n(T\omega) = x_{n+1}(\omega)$. Так как $x_n(\omega) = x_0(T^n\omega)$, то любое утверждение относительно случайных величин x_n может быть сформулировано как утвер-

---

[1]) В конце книги помещен указатель примеров.

[2]) Эти характеристики Ω и $\mathscr{F}$ можно взять в качестве определений. Из общей теории произведений мер нам нужна теорема существования на $\mathscr{F}$ единственной вероятностной меры P, удовлетворяющей условию (1.1). Это утверждение следует также из теоремы существования Колмогорова (или Даниеля — Колмогорова). Элементарное доказательство см. в § 3.

ждение относительно x_0 и T. Если множество A — цилиндр, то очевидно, что $T^{-1}A$ — также цилиндр и, следовательно, принадлежит $\mathcal{F}$, причем $P(T^{-1}A) = P(A)$. Тогда T — измеримое сохраняющее меру преобразование, что вытекает из следующего общего результата.

Теорема 1.1. *Пусть $\mathcal{F}_0$ — поле, порождающее $\mathcal{F}$. Если $T^{-1}A \in \mathcal{F}$ и $P(T^{-1}A) = P(A)$ для всех A из $\mathcal{F}_0$, то T — сохраняющее меру преобразование.*

Доказательство. Совокупность $\mathcal{G}$ множеств A из $\mathcal{F}$, для которых $T^{-1}A \in \mathcal{F}$ и $P(T^{-1}A) = P(A)$, является монотонным классом и содержит поле $\mathcal{F}_0$, а следовательно, совпадает с $\mathcal{F}$ (см. Халмош [2]).

Пример 1.1 дает математическую модель бесконечной в обе стороны последовательности испытаний Бернулли. Представьте себе простой эксперимент с конечным пространством состояний ρ и вероятностью p_i исхода i. Пример является моделью мысленного эксперимента, состоящего в том, что производятся независимые повторения простого эксперимента, образующие бесконечную в обе стороны последовательность. Пусть каждый день производится один простой эксперимент. Точка $\omega = (\ldots, \omega_{-1}, \omega_0, \omega_1, \ldots)$ пространства Ω полностью определяет исход большого эксперимента, а ее компонента ω_n — исход простого эксперимента, произведенного в n-й день.

Здесь может возникнуть некоторая путаница. Мы не предполагаем, что знание исхода эксперимента, скажем, 5-го дня помогает предсказать исход эксперимента 6-го дня. Но исход $x_5(\omega)$ не определен, пока не определен исход большого эксперимента ω, а с определением ω полностью определен и $x_6(\omega)$. Создается впечатление, что случайность исчезает. Иными словами, все координаты точки ω определяются, так сказать, одновременно. Как же тогда они могут служить моделью экспериментов, совершаемых последовательно во времени?

При взгляде на эту модель полезна следующая аллегория. Тихе[1]) выбирает точку ω из пространства Ω в соответствии с распределением вероятностей P. Она совершает это до начала отсчета времени и затем раскрывает экспериментатору одну за другой координаты ω_n — по одной каждый

[1]) Богиня случая; см. Грейвс [1]. Ее появление в этой книге каждый раз указывает на эвристический оттенок рассуждений.

день. Динамические аспекты модели удобно изучать с помощью преобразования T, так как оно связано с ходом времени. (Сама математика, к счастью, не нуждается в подобных интерпретациях.)

Замечание об обратимости. Применение теоремы 1.1 к обратному точечному отображению T^{-1} показывает, что преобразование примера 1.1 обратимо. В некоторых случаях обратное точечное отображение вполне определено (т. е. прямое преобразование взаимно однозначно и имеет областью значений все пространство Ω), но не обязательно измеримо. Это имеет место, например, для преобразования T, определенного на вероятностном пространстве $(\Omega, \mathcal{F}', P')$, где T и Ω таковы, как в примере 1.1, $\mathcal{F}'$ есть σ-поле, порожденное $x_0, x_1, x_2, \ldots$, а P' — сужение P на $\mathcal{F}'$. Измеримость T^{-1} требует некоторого обоснования. Однако если это установлено, то из сохранения меры преобразованием T следует непосредственно сохранение меры преобразованием T^{-1}.

Пример 1.2. Пусть Ω и $\mathcal{F}$ те же, что в первом примере, но пусть теперь P — *любая* мера, которая сохраняется при определенном выше преобразовании T. Сказать, что преобразование T сохраняет меру P, равносильно (в силу теоремы 1.1) утверждению независимости $P\{\omega: (x_n(\omega), \ldots, x_{n+k-1}(\omega)) \in E\}$ от n, что как раз совпадает с определением стационарности стохастического процесса $\{\ldots, x_{-1}, x_0, x_1, \ldots\}$, образованного координатными случайными величинами. Преобразование T называется *сдвигом*, связанным с этим процессом.

Так как совокупность конечных объединений непересекающихся тонких цилиндров образует поле, порождающее $\mathcal{F}$, то мера P на $\mathcal{F}$ однозначно определяется своими значениями

$$p_k(i_1, \ldots, i_k) = P\{\omega: x_n(\omega) = i_1, \ldots, x_{n+k-1}(\omega) = i_k\} \quad (1.2)$$

на этих цилиндрах. Если функции p_k на последовательностях длины k элементов из ρ определяются формулой (1.2), то

$$\begin{gathered} p_k(i_1, \ldots, i_k) \geqslant 0, \\ \sum_i p_{k+1}(i_1, \ldots, i_k, i) = p_k(i_1, \ldots, i_k), \\ \sum_i p_1(i) = 1, \end{gathered} \quad (1.3)$$

и так как преобразование T сохраняет меру, то

$$p_k(i_1, \ldots, i_k) = \sum_i p_{k+1}(i, i_1, \ldots, i_k). \tag{1.4}$$

Обратно, существует ровно одна вероятностная мера P на $\mathcal{F}$, соответствующая данному множеству функций p_k, удовлетворяющих соотношениям (1.3) и (1.4), для которой выполняется (1.2) и которая сохраняется при преобразовании T[1]). Если $p_k(i_1, \ldots, i_k) = p_{i_1} p_{i_2} \ldots p_{i_k}$, где p_i — неотрицательные числа, в сумме дающие 1, то имеют место соотношения (1.3) и (1.4). Мы снова приходим к примеру 1.1, который будем называть *сдвигом Бернулли*[2]). В § 3 детально разбираются другие специальные случаи. Заметим, что преобразование T как точечное преобразование, определенное на пространстве Ω, — одно и то же в примерах 1.1 и 1.2. Однако преобразование, сохраняющее меру, является точечным преобразованием с σ-полем, относительно которого оно измеримо и сохраняет меру.

Пример 1.2 дает модель испытаний, не обязательно независимых, которые производятся в постоянных условиях, — идея, стоящая за стационарностью, т. е. за требованием сохранения меры P преобразованием T. Так как стационарные процессы имеют широкое распространение, а любое свойство такого процесса является по существу свойством соответствующего сдвига, то имеет смысл изучать сам сдвиг. Наилучшим способом изучения сдвига является изучение сохраняющих меру преобразований, ибо это предоставит в наше распоряжение множество примеров, большая часть которых значительно проще двух вышеприведенных. Разберем здесь еще четыре примера.

Пример 1.3. Пусть Ω — пространство из пяти точек $\{a, b, c, d, e\}$, $\mathcal{F}$ — множество всех подмножеств из Ω и T — подстановка, равная произведению двух циклов: $T = (a, b, c)(d, e)$. Если T сохраняет P, то точки в пределах циклов должны быть равновероятны относительно P.

Пример 1.4. Для Ω и $\mathcal{F}$ из предыдущего примера возьмем в качестве T циклическую подстановку $T = (a, b, c, d, e)$. Все пять точек должны быть равновероятны.

---

[1]) Это частный случай теоремы существования Колмогорова; доказательство дано в § 3.

[2]) В отечественной литературе принят термин *автоморфизм Бернулли. — Прим. ред.*

Пример 1.5. Пусть Ω — единичная окружность на комплексной плоскости, а $\mathcal{F}$ состоит из обычных борелевских подмножеств окружности Ω ($\mathcal{F}$ есть σ-поле, порожденное дугами окружности), и пусть P — мера Лебега на окружности, нормированная таким образом, что $P(\Omega)=1$. Пусть $T\omega = c\omega$, где c — фиксированный элемент пространства Ω. Так как T есть просто поворот окружности на угол, равный $\arg c$, то T сохраняет P. Далее мы увидим, что свойства преобразования T существенно зависят от того, является ли c корнем из единицы.

Пример 1.6. Пусть $\mathcal{F}$ состоит из борелевских подмножеств полуоткрытого единичного интервала $\Omega=[0, 1)$, P — мера Лебега и $T\omega = 2\omega \pmod 1$ (т. е. $T\omega$ равняется 2ω на $[0, 1/2)$ и $2\omega-1$ на $[1/2, 1)$). Преобразование T тесно связано с диадическим (по основанию 2) разложением точек единичного интервала; действительно, если $f(\omega)=0$ для $\omega<1/2$ и $f(\omega)=1$ для $\omega\geqslant 1/2$, то $f(T^{n-1}\omega)$ является n-м знаком диадического разложения $\omega=\sum_{n=1}^{\infty} f(T^{n-1}\omega)/2^n$ (если ω — двоично-рациональное число, то разложение конечно). Следовательно, если разложение ω имеет вид $\omega=0, \omega_1\omega_2 \ldots$, то $T\omega=0, \omega_2\omega_3 \ldots$. Применение теоремы 1.1 показывает, что преобразование T сохраняет меру. В то время как все предыдущие преобразования были обратимы, это преобразование, которое мы будем называть *диадическим преобразованием*, необратимо.

Мы привели достаточно примеров преобразований, которые помогут иллюстрировать наши дальнейшие рассмотрения. Некоторое количество дополнительных примеров приводится в § 3.

Эргодичность

Нас будет интересовать вопрос: какие преобразования T обладают тем свойством, что для почти всех ω *траектория* $\{\omega, T\omega, T^2\omega, \ldots\}$[1]) точки ω точно воспроизводит само пространство Ω? В примере 1.4 каждая траектория воспроизводит Ω в том простом смысле, что она как множество

---

[1]) Мы считаем эту последовательность траекторией точки ω, даже если преобразование T обратимо (хотя правильнее было бы называть ее полутраекторией, а траекторией — последовательность $\{\ldots, T^{-1}\omega, \omega, T\omega, \ldots\}$).

совпадает с Ω, но это неверно для примера 1.3, где множество $A=\{a, b, c\}$ переводится преобразованием T в себя, так же как и дополнительное множество $\{d, e\}$: $T^{-1}A=A$. Заметим, что в примере 1.4 каждая траектория не только совпадает как множество с Ω, но и содержит элементы из Ω в правильной пропорции: асимптотическая относительная частота, с которой элемент a (или $b, \ldots$) встречается в траектории, в точности равна $P(a)=1/5$ (или $P(b)=1/5, \ldots$).

Замечательно следующее: если для некоторого сохраняющего меру преобразования T не существует множеств A таких, что $T^{-1}A=A$, за исключением „неинтересных" множеств $A=\varnothing$ и $A=\Omega$, то траектории воспроизводят пространство Ω в том смысле, что для каждого множества A траектории почти всех точек ω попадают в A с асимптотической относительной частотой $P(A)$:

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1} I_A(T^k\omega)=P(A) \text{ п. в.}$$[1].

Это центральный факт в эргодической теории. Он будет описан и проиллюстрирован в этом параграфе, а доказан в следующем.

Назовем множество[2] A *инвариантным* (относительно преобразования T), если $T^{-1}A=A$; в случае когда T обратимо, это требование эквивалентно требованию $TA=A$. Назовем преобразование T *эргодическим*[3]), если каждое инвариантное относительно него множество тривиально, т. е. имеет меру 0 или 1. Следовательно, преобразование примера 1.4 эргодично, а преобразование примера 1.3 не эргодично (если только вся масса не сосредоточена в одном из циклов). Мы убедимся далее, что если преобразование T эргодично, то траектории обладают только что описанным свойством.

По техническим соображениям удобно несколько видоизменить определение инвариантности: назовем множество A инвариантным относительно преобразования T, если

---

[1]) I_A — характеристическая функция множества A; таким образом, $\sum_{k=0}^{n-1} I_A(T^k\omega)$ есть число элементов последовательности $\{\omega, T\omega, \ldots, T^{n-1}\omega\}$, лежащих в A. Символ „п. в." (почти всюду) означает „с точностью до множества меры 0".

[2]) Относительно любого множества всюду, где не оговорено противное, предполагается, что оно принадлежит полю $\mathcal{F}$.

[3]) *Эргодическое* преобразование называют также *метрически транзитивным* или *неразложимым*.

$P(A+T^{-1}A)=0$[1]. Например, если точки d и e в примере 1.3 имеют нулевую массу, то множество $A=\{a, b, c, d\}$ инвариантно в новом смысле, в то время как в старом смысле оно не инвариантно. Однако для любого множества A множество $B=\overline{\lim\limits_n}\, T^{-n}A$ инвариантно в старом смысле, и в случае когда A инвариантно в новом смысле, множества A и B имеют одинаковую меру. Поэтому, если существует нетривиальное множество, инвариантное в новом смысле, то существует и нетривиальное множество, инвариантное в старом смысле, так что определение эргодичности не меняется. Множество, инвариантное в старом смысле, мы будем называть *строго* инвариантным. Для доказательства эргодичности достаточно убедиться, что строго инвариантные множества имеют меру 0 или 1.

Замечание. Так как преобразование T сохраняет меру, то

$$P(A-A\cap T^{-1}A)=P(A)-P(A\cap T^{-1}A)=$$
$$=P(T^{-1}A)-P(A\cap T^{-1}A)=P(T^{-1}A-A\cap T^{-1}A)$$

и, следовательно,

$$P(A+T^{-1}A)=2P(A-T^{-1}A)=2P(T^{-1}A-A).$$

Поэтому множество A инвариантно в том и только том случае, если одна из мер $P(A-T^{-1}A)$ или $P(T^{-1}A-A)$ равна 0. В частности, если T эргодично, то ни одно из соотношений $A\subset T^{-1}A$, $T^{-1}A\subset A$ не может выполняться для нетривиального A: не будучи в состоянии оставить нетривиальное множество фиксированным, эргодическое преобразование не может ни „расширить“, ни „сузить“ его.

Очевидно, что если преобразование T обратимо, то оно эргодично тогда и только тогда, когда эргодично преобразование T^{-1}.

Эргодичность вращений

Если в примере 1.5 число c, определяющее поворот, равно -1, то множество, состоящее из первого и третьего квадрантов, является нетривиальным инвариантным множеством, и, следовательно, преобразование T неэргодично.

[1]) Знак $+$ означает здесь симметрическую разность: $A+B=(A-B)\cup(B-A)$. Мы будем часто пользоваться соотношениями $A+B\subset(A+C)+(C+B)$ и $P(A+B)\leqslant P(A+C)+P(C+B)$.

Аналогичное построение показывает, что T неэргодично, если c — любой корень из единицы.

Покажем, что если c не является корнем из единицы, преобразование T эргодично. Пусть $e_n(\omega)=\omega^n$ — круговые функции; ряд Фурье для характеристической функции I_A есть $I_A(\omega)\sim\sum\limits_n a_n e_n(\omega)$. Так как $e_n(T\omega)=c^n e_n(\omega)$, то с помощью замены переменной получаем [1])

$$a_n=\int\limits_A e_{-n}\,dP=c^{-n}\int\limits_{T^{-1}A} e_{-n}\,dP,$$

откуда $I_{T^{-1}A}(\omega)\sim\sum\limits_n c^n a_n e_n(\omega)$. Если A — инвариантное множество, то $I_A(\omega)$ и $I_{T^{-1}A}(\omega)$ почти всюду равны и, следовательно, имеют одинаковые коэффициенты Фурье: $a_n=c^n a_n$ для всех n. Если c не является корнем из единицы, то $a_n=0$ для всех $n\neq 0$, и в силу теоремы единственности для коэффициентов Фурье $I_A(\omega)$ равно почти всюду некоторой константе, так что мера $P(A)$ должна быть равна либо нулю, либо единице. Следовательно, преобразование T эргодично.

Старый теоретико-числовой результат, принадлежащий Якоби, состоит в том, что если c не является корнем из единицы, то траектория каждой точки ω всюду плотна на единичной окружности Ω, что представляет собой еще одно простое условие эргодичности. (Заметим, что в данном случае траектория не совпадает как множество с пространством Ω, ибо является счетным множеством, в то время как Ω несчетно.) Для доказательства результата Якоби, очевидно, достаточно показать, что траектория $\{1, c, c^2, \ldots\}$ точки 1 всюду плотна. Но если c — не корень из единицы, то все точки этой траектории различны и, следовательно,

---

[1]) Функция $f(\omega)$ интегрируема на A в том и только том случае, если функция $f(T\omega)$ интегрируема на $T^{-1}A$, причем тогда $\int\limits_{T^{-1}A} f(T\omega)\,P(d\omega)=\int\limits_A f(\omega)\,P(d\omega)$. При доказательстве достаточно рассмотреть случай $A=\Omega$, общий случай получится при замене f на $I_A f$. Если f — характеристическая функция, то $\int f(T\omega)\,P(d\omega)=\int f(\omega)\,P(d\omega)$, так как T — сохраняющее меру преобразование; эта формула легко получается для простой функции, а затем, с помощью аппроксимации, для любой функции f.

в силу компактности имеют предельную точку ω_0. Поэтому для любого положительного ε существуют различные точки c^n и c^{n+k}, отстоящие от ω_0 не больше чем на $\varepsilon/2$ (расстояние измеряется по дуге) и, следовательно, не больше чем на ε друг от друга. Так как расстояние от точки c^{n+lk} до точки $c^{n+(l+1)k}$ равно расстоянию от c^n до c^{n+k}, то очевидно, что для некоторого m точки $c^n, c^{n+k}, \ldots, c^{n+mk}$ образуют такую цепь вдоль всей окружности, что расстояние между любыми соседними точками меньше ε. Таким образом, любая точка окружности имеет в своей ε-окрестности точку траектории $\{1, c, c^2, \ldots\}$, и траектория всюду плотна в силу произвольности выбора ε.

Используем результат Якоби для того, чтобы дать второе доказательство эргодичности вращения T (по-прежнему предполагая, что c не является корнем из единицы). Это доказательство не использует теорему единственности для коэффициентов Фурье. Пусть A — строго инвариантное множество положительной меры, $P(A) > 0$. Докажем, что $P(A) = 1$.

Сначала покажем, что для любого ε из единичного интервала $(0 < \varepsilon < 1)$ существует невырожденная дуга I длины, не превышающей ε, такая, что $P(A \cap I) \geqslant (1-\varepsilon) P(I)$. Действительно, по определению меры Лебега на окружности множество A может быть покрыто последовательностью дуг $I_1, I_2, \ldots$ так, что $P(A)/(1-\varepsilon) \geqslant \sum\limits_n P(I_n)$ (мы предположили, что $P(A) > 0$); в качестве членов этой последовательности можно выбрать непересекающиеся дуги длины меньшей, чем ε[1]). Так как $\sum\limits_n P(A \cap I_n) = P(A) \geqslant (1-\varepsilon) \sum\limits_n P(I_n)$, то для некоторого значения n должно выполняться неравенство $P(A \cap I_n) \geqslant (1-\varepsilon) P(I_n)$. Возьмем $I = I_n$.

Далее, так как множество A инвариантно, а преобразование T обратимо и сохраняет меру P, получаем $P(A \cap T^n I) \geqslant (1-\varepsilon) P(T^n I)$. Если $n_1, \ldots, n_k$ — натуральные числа, для которых множества $T^{n_1}I, \ldots, T^{n_k}I$ не пересекаются, то

$$P(A) \geqslant \sum_{i=1}^{k} P(A \cap T^{n_i} I) \geqslant (1-\varepsilon) P\left(\bigcup_{i=1}^{k} T^{n_i} I\right).$$

Если траектория $\{1, c, c^2, \ldots\}$ всюду плотна, то траектория любой из концевых точек дуги I также всюду плотна. Так

[1]) Если A содержит интервал, то этого сделать нельзя, но тогда доказываемое утверждение очевидно. — *Прим. ред.*

как $P(I) \leqslant \varepsilon$, то существуют, очевидно, такие натуральные числа $n_1, \ldots, n_k$, что множества $T^{n_1}I, \ldots, T^{n_k}I$, не пересекаясь, примыкают друг к другу так тесно, что покрывают всю окружность, кроме, быть может, множества меры 2ε, т. е. $P\left(\bigcup_{i=1}^{k} T^{n_i}I\right) \geqslant 1-2\varepsilon$. Итак,

$$P(A) \geqslant (1-\varepsilon)(1-2\varepsilon),$$

а так как ε произвольно, то $P(A)=1$. Эргодичность T доказана.

Строго инвариантные множества являются объединениями полных траекторий (множеств вида $(\ldots, T^{-1}\omega, \omega, T\omega, \ldots)$), что вызывает иллюзию возможности построения нетривиального инвариантного множества из „половин" полных траекторий. Такое построение возможно, когда c — корень из единицы, но неприменимо в противном случае: мы видели, что если c не является корнем из единицы, то любое строго инвариантное борелевское множество имеет меру 0 или 1, а тогда это верно и для любого строго инвариантного лебеговского множества [1]).

Эргодичность диадического преобразования

Рассмотрим преобразование $T\omega = 2\omega \pmod 1$ (пример 1.6). Если $\omega = 0, \omega_1\omega_2 \ldots$ и $\omega' = \omega + 1/2 \pmod 1$, то $\omega' = 0, \omega_1'\omega_2\omega_3 \ldots$, где $\omega_1' = 1 - \omega_1$. Если $A = T^{-1}A$, то $\omega \in A$ эквивалентно $T\omega \in A$ и $\omega' \in A$ эквивалентно $T\omega' \in A$. Так как $T\omega = T\omega' = 0, \omega_2\omega_3 \ldots$, то $\omega \in A$ в том и только том случае, если $\omega' \in A$. Следовательно, если $E = [0, 1/2)$, то $A \cap E^c$ [2]) представляет собой сдвинутое на $1/2$ вправо множество $A \cap E$. Тогда эти два множества имеют одинаковую меру (здесь P — мера Лебега) и $P(A) = 2P(A \cap E) = P(A \cap E)/P(E)$. Таким образом, A и E независимы. Можно показать, что это верно и в случае когда E — любой диадический интервал или объединение

---

[1]) В случае когда c не является корнем из единицы, существуют строго инвариантные неизмеримые множества; такие множества автоматически имеют внутреннюю меру 0 и внешнюю меру 1. Упорядочим по включению семейство строго инвариантных множеств, не содержащих пары точек ω_1 и ω_2, эквивалентных в том смысле, что ω_1/ω_2 — корень из единицы. В силу леммы Цорна это семейство содержит максимальный элемент A. Любая точка ω эквивалентна некоторой точке множества A, так как иначе A можно было бы расширить, присоединив к нему полную траекторию ω. Из сказанного следует, что A неизмеримо.

[2]) E^c обозначает дополнение множества E.

непересекающихся диадических интервалов. Для заданного положительного ε выбираем объединение E так, чтобы $P(A+E)<\varepsilon$. Тогда $|P(A)-P(E)|<\varepsilon$ и $|P(A)-P(A)P(E)|=$ $=|P(A)-P(A\cap E)|<\varepsilon$, так что $|P(A)=P^2(A)|<2\varepsilon$. В силу произвольности ε имеем $P(A)=P^2(A)$, откуда $P(A)$ равняется либо 0, либо 1. Следовательно, преобразование T эргодично.

Перемешивание

Преобразование примера 1.1 оказывается эргодическим, но мы покажем, что оно обладает даже более сильным свойством. Сохраняющее меру преобразование T называется *перемешивающим*, если равенство

$$\lim_{n\to\infty} P\left(A\cap T^{-n}B\right)=P(A)P(B) \tag{1.5}$$

справедливо для любой пары множеств A и B [1]. Если T обратимо, то эквивалентным условием перемешивания является условие (1.5), в котором $T^{-n}B$ заменено на T^nB. Так как (1.5) эквивалентно условию $P\left(T^{-n}B\mid A\right)-P\left(T^{-n}B\right)\to 0$, то свойство перемешивания имеет смысл и без предположения инвариантности меры.

Если множество B из (1.5) инвариантно, то $P(A\cap B)=$ $=P(A)P(B)$, и если взять $A=B$, то $P(B)$ равняется 0 или 1. Следовательно, *перемешивание влечет эргодичность.* В большом числе случаев свойство перемешивания можно выявить с помощью следующего результата.

Теорема 1.2. *Пусть $\mathcal{F}_0$ — поле, порождающее $\mathcal{F}$. Если* (1.5) *выполняется для любых A и B из $\mathcal{F}_0$, то преобразование T перемешивающее.*

Доказательство. Для данных множеств A и B из $\mathcal{F}$ и некоторого положительного ε выберем множества A_0 и B_0 из $\mathcal{F}_0$ так, чтобы мера множеств $A+A_0$, $B+B_0$ была меньше ε. Тогда $T^{-n}B+T^{-n}B_0=T^{-n}(B+B_0)$ имеет меру меньше ε и, следовательно, мера множества $\left(A\cap T^{-n}B\right)+$ $+\left(A_0\cap T^{-n}B_0\right)$ меньше 2ε для всех n. Таким образом,

---

[1] Это свойство иногда называют *сильным перемешиванием*, чтобы отличить его от свойства *слабого перемешивания*, для которого требуется только выполнение условия $n^{-1}\sum_{k=0}^{n-1}|P(A\cap T^{-k}B)-P(A)P(B)|\to 0$. См. Халмош [3].

$P(A \cap T^{-n}B)$ отличается от $P(A_0 \cap T^{-n}B_0)$ не больше, чем на 2ε, и, следовательно, ее верхний и нижний пределы не больше, чем на 2ε отличаются от $\lim\limits_{n\to\infty} P(A_0 \cap T^{-n}B_0) = P(A_0 \cap B_0)$, что в свою очередь не больше, чем на 2ε отличается от $P(A \cap B)$. Результат следует из произвольности выбора ε.

Если множества A и B — цилиндры из примера 1.1, то A и $T^{-n}B$ — цилиндры, зависящие от непересекающихся множеств координат, если n достаточно велико. Тогда $P(A \cap T^{-n}B) = P(A)P(B)$ для больших n, так что (1.5) выполняется для всех множеств A и B из поля цилиндров $\mathcal{F}_0$. В силу теоремы 1.2 T — перемешивающее и, следовательно, эргодическое преобразование.

Преобразование примера 1.3, разумеется, не обладает свойством перемешивания, ибо оно не эргодично. Преобразование примера 1.4 эргодическое, но не перемешивающее (возьмем $A = B = \{a\}$). Более интересный пример эргодического, но не перемешивающего преобразования дает вращение (пример 1.5) при c, не равном корню из единицы. Пусть $A = B$ — верхняя полуокружность; так как траектория $\{c^n\}$ всюду плотна, множества A и $T^{-n}A$ почти совпадают для бесконечно большого числа значений n, так что условие (1.5) не выполняется. Таким образом, перемешивание по сравнению с эргодичностью — свойство более сильное.

С помощью теоремы 1.2 можно убедиться, что диадическое преобразование (пример 1.6) не только эргодично, но и обладает свойством перемешивания.

Формулировка эргодической теоремы

Займемся теперь основными следствиями из эргодичности. Функция $g(\omega)$ (которая предполагается измеримой относительно $\mathcal{F}$) называется *инвариантной*, если $g(T\omega) = g(\omega)$ почти всюду. Множество инвариантно в том и только том случае, если инвариантна его характеристическая функция. Если g — инвариантная нетривиальная (т. е. не равная почти всюду константе) функция, то для некоторого α мера инвариантного множества $\{\omega\colon g(\omega) \leqslant \alpha\}$ заключена строго между 0 и 1. Таким образом, преобразование T эргодично в том и только том случае, если всякая инвариантная функция равна константе почти всюду. Мы можем теперь сформулировать эргодическую теорему.

Теорема 1.3. *Если f интегрируема, то существует такая интегрируемая инвариантная функция $\hat{f}$, что $E\{\hat{f}\} = E\{f\}$*[1]) *и*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f(T^k\omega) = \hat{f}(\omega) \text{ п. в.} \tag{1.6}$$

Если преобразование T эргодично, то $\hat{f}(\omega) = E\{f\}$ п. в.

Мы дадим сейчас доказательство менее сильного утверждения для весьма специального случая: мы покажем, что если T — перемешивающее преобразование и f — характеристическая функция множества A, то $\frac{1}{n}\sum_{k=0}^{n-1} f(T^k\omega)$ сходится по вероятности к $E\{f\} = P(A)$. Если

$$c_{i,k} = E\{(f(T^i\omega) - P(A))(f(T^k\omega) - P(A))\} =$$
$$= P(T^{-i}A \cap T^{-k}A) - P(A)P(A),$$

то, так как T сохраняет меру, $c_{i,k} = \rho_{|k-i|}$, где $\rho_n = P(A \cap T^{-n}A) - P(A)P(A) \to 0$ при $n \to \infty$ в силу предположения о перемешивающем свойстве преобразования T. По теореме об арифметических средних сходящейся последовательности имеем

$$E\left\{\left[\frac{1}{n}\sum_{k=0}^{n-1} f(T^k\omega) - P(A)\right]^2\right\} = \frac{1}{n^2}\sum_{i=0}^{n-1}\sum_{k=0}^{n-1} c_{i,k} =$$
$$= \frac{1}{n}\rho_0 + \frac{2}{n^2}\sum_{k=1}^{n-1}(n-k)\rho_k \leqslant \frac{1}{n}|\rho_0| + \frac{2}{n}\sum_{k=1}^{n-1}|\rho_k| \to 0.$$

Сходимость по вероятности $\frac{1}{n}\sum_{k=0}^{n-1} f(T^k\omega)$ к $P(A)$ следует теперь из неравенства Чебышева. Доказательство для общего случая, непохожее на приведенное, дано в следующем параграфе.

Если предел (1.6) существует почти всюду, то, очевидно, предельная функция $\hat{f}(\omega)$ инвариантна и в случае эргодичности преобразования равна константе почти всюду. Так как $E\{\hat{f}\} = E\{f\}$, эта константа есть $E\{f\}$. Если T эргодично, то ω-множество, для которого существует предел (1.6), будучи инвариантным, имеет меру 0 или 1, но доказательство того, что мера его равна 1, затруднительно даже в этом случае.

---

[1]) Символы $E\{f\}$, $\int f(\omega)\,P(d\omega)$ и $\int f\,dP$ взаимозаменяемы.

Если T неэргодично, то предельная функция $\hat{f}$ не обязательно постоянна, так как если f — нетривиальная инвариантная функция (скажем, характеристическая функция нетривиального инвариантного множества), то средние арифметические в соотношении (1.6) равны самой f, так что $\hat{f} = f$ не равна константе.

В § 10 (пример 10.2) мы покажем, что $\hat{f}$ является значением условного математического ожидания (а именно значением условного математического ожидания функции f относительно σ-поля инвариантных множеств).

Следствия из эргодической теоремы

Если преобразование T эргодично, то, положив в соотношении (1.6) $f = I_A$, имеем

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} I_A(T^k\omega) = P(A) \text{ п. в.} \tag{1.7}$$

Именно в этом смысле траектории эргодического преобразования воспроизводят пространство Ω.

Применение эргодической теоремы к сдвигу Бернулли (пример 1.1) приводит к усиленному закону больших чисел для испытаний Бернулли. Действительно, пусть $f(\omega)$ принимает значения 1 или 0 в зависимости от того, является $x_1(\omega)$ элементом i множества ρ или нет (f — характеристическая функция цилиндра $\{\omega\colon\ x_1(\omega) = i\}$). Тогда $\sum_{k=0}^{n-1} f(T^k\omega)$ указывает, сколько раз встречается элемент i среди $x_1(\omega), \ldots, x_n(\omega)$, так что предел (если он существует) выражения $\frac{1}{n} \sum_{k=0}^{n-1} f(T^k\omega)$ представляет собой асимптотическую относительную частоту, с которой исход i встречается в той части бесконечной последовательности испытаний ω, которая соответствует положительным точкам временной оси. Согласно эргодической теореме, этот предел существует и равен $E\{f\} = p_i$ почти всюду, т. е. имеет место усиленный закон больших чисел.

Придавая $f(\omega)$ значения 1 или 0 в зависимости от того, выполняются или нет равенства $x_1(\omega) = i$, $x_2(\omega) = j$, видим, что асимптотическая относительная частота исходов i и j в последовательных испытаниях равна $p_i p_j$ почти всюду. Заметим, что эргодическая теорема для сдвига Бернулли

сильнее, чем подобные утверждения, так как в качестве f может быть взята некоторая сложная функция *всех* координатных переменных.

Из применения эргодической теоремы к диадическому преобразованию (пример 1.6) следует теорема Бореля о нормальных числах. Число ω в единичном интервале нормально (по основанию 2), если относительная частота единиц среди первых n знаков диадического разложения ω сходится к 1/2. Из эргодической теоремы с $f(\omega)=0$ для $\omega<1/2$ и $f(\omega)=1$ для $\omega\geqslant 1/2$ следует, что почти все числа нормальны.

Если же к этому примеру применить эргодическую теорему с $f(\omega)=\omega$, то мы увидим, что $n^{-1}\sum_{k=0}^{n-1}\{2^k\omega\}\to 1/2$ почти всюду (символ $\{x\}$ означает дробную часть действительного числа x).

Рассмотрим теперь вращение окружности, когда c не является корнем из единицы. В этом случае преобразование T эргодично, и в силу эргодической теоремы для почти всех ω траектория $\{\omega, c\omega, c^2\omega, \ldots\}$ точки ω попадает в любую заданную дугу I с „правильной“ асимптотической относительной частотой, а именно с частотой $P(I)$. Пусть $I_1, I_2, \ldots$ — все дуги, концевые точки которых принадлежат некоторому фиксированному счетному всюду плотному подмножеству из Ω, скажем, множеству всех корней из единицы. Тогда для почти всех ω траектория ω попадает в каждую дугу I_n с соответствующей частотой; нетрудно показать, что траектория любой такой точки ω *равномерно распределена* в том смысле, что она попадает во *всякую* дугу окружности с правильной асимптотической относительной частотой[1]). Таким образом, траектории почти всех точек равномерно распределены, а так как каждая траектория получается из другой поворотом, то каждая траектория равномерно распределена. Разумеется, равномерно распределенная траектория всюду плотна, что снова приводит нас к теореме Якоби.

Каждое из этих следствий эргодической теоремы (за исключением, конечно, основного результата (1.7)) было доказано с помощью специальных методов раньше, чем была обнаружена сама эргодическая теорема.

---

[1]) Можно даже показать (используя теорию слабой сходимости), что равномерно распределенная траектория попадает в произвольное борелевское множество A с предельной относительной частотой $P(A)$, если его граница имеет меру 0. Однако для каждого ω существует некоторое множество, в которое траектория попадает с неправильной частотой, например сама траектория точки ω.

Критерии эргодичности

Рассматривая свойства эргодичности и перемешивания преобразований из различных примеров, мы избегали обсуждения с этой точки зрения общего сдвига примера 1.2. Дело в том, что не существует эффективного критерия для решения вопроса о том, является ли данный сдвиг эргодическим или перемешивающим; вопрос этот должен решаться отдельно для каждого такого случая или класса случаев.

Приведем построение, показывающее, что сдвиг примера 1.2 является весьма общим. Пусть $\tilde{T}$ — обратимое сохраняющее меру преобразование, определенное на вероятностном пространстве $(\tilde{\Omega}, \tilde{\mathscr{F}}, \tilde{P})$, и $\tilde{A}$ — некоторое множество, принадлежащее полю $\tilde{\mathscr{F}}$. Пусть $(\Omega, \mathscr{F})$ — произведение пространств из примера 1.2 с $\rho = \{0, 1\}$. Определим отображение φ пространства $\tilde{\Omega}$ в Ω с помощью характеристической функции $(\varphi(\tilde{\omega}))_n = I_{\tilde{A}}(\tilde{T}^n\tilde{\omega})$. Если $P = \tilde{P}\varphi^{-1}$, то сдвиг T на пространстве Ω сохраняет меру P. Таким способом могут быть построены сдвиги более или менее произвольной сложности.

Следующая теорема иногда используется для установления эргодичности сдвигов и некоторых других преобразований (применения даны в § 3). Если соотношение (1.7) выполняется, будем говорить, что A — *респект*[1]) для T; таким образом, всякое множество A является респектом эргодического преобразования T. Если B — респект для T, то

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} I_A(\omega) I_B(T^k\omega) = I_A(\omega) P(B) \text{ п. в.}$$

для каждого множества A. Интегрируя под знаком предела, получаем

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} P\left(A \cap T^{-k}B\right) = P(A) P(B). \tag{1.8}$$

Итак, каково бы ни было множество A, (1.8) выполняется, если только B является респектом для T. С другой стороны, если (1.8) имеет место для всех A и B, легко показать, что мера любого инвариантного множества равна 0 или 1 и, следовательно, T эргодично. Наконец, если соотношение (1.8)

[1]) В оригинале «T respects A». Можно показать, что если A — респект для T, то A не зависит от σ-поля T-инвариантных множеств (см. пример 10.2). — *Прим. ред.*

выполняется для всех множеств A и B из поля $\mathcal{F}_0$, порождающего $\mathcal{F}$, то оно выполняется и для всех множеств A и B из $\mathcal{F}$; этот результат можно получить с помощью рассуждений, аналогичных тем, которые были использованы при доказательстве теоремы 1.2. Мы имеем, таким образом, следующий критерий.

Теорема 1.4. *Если T — эргодическое преобразование, то каждое множество A является респектом для T и соотношение* (1.8) *выполняется для всех A и B. Пусть $\mathcal{F}_0$ — поле, порождающее $\mathcal{F}$. Если каждое множество A из $\mathcal{F}_0$ является респектом для T или если соотношение* (1.8) *выполняется для всех A и B из $\mathcal{F}_0$, то $\mathcal{F}_0$ — эргодическое преобразование.*

Сопоставление соотношений (1.5) и (1.8) показывает, что перемешивание в сравнении с эргодичностью является более сильным свойством.

Даже если преобразование T не эргодично, то

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} I_B(T^k\omega) = \hat{I}_B(\omega) \quad \text{п. в.}$$

Интегрируя по A, получаем

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} P\left(A \cap T^{-k}B\right) = E\{I_A \hat{I}_B\}.$$

Правую часть этого равенства можно записать в другой форме, симметричной относительно A и B. Действительно, функция $(I_A\hat{I}_B)^\wedge$, соответствующая функции $I_A\hat{I}_B$, имеет почти всюду значение

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} I_A(T^k\omega)\,\hat{I}_B(T^k\omega) =$$

$$= \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} I_A(T^k\omega)\,\hat{I}_B(\omega) = \hat{I}_A(\omega)\,\hat{I}_B(\omega),$$

где первое равенство выполняется почти всюду в силу инвариантности $\hat{I}_B$. Поэтому, так как функции $I_A\hat{I}_B$ и $(I_A\hat{I}_B)^\wedge$ имеют одинаковый интеграл, то

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} P\left(A \cap T^{-k}B\right) = E\{\hat{I}_A \hat{I}_B\}. \tag{1.9}$$

Предел в соотношении (1.8) всегда существует.

Более сложный сдвиг* [1]

Наша заключительная иллюстрация развивает идею примера 1.2.

Пример 1.7. Пусть Ω состоит из бесконечных в обе стороны последовательностей $\omega = (\ldots, \omega_{-1}, \omega_0, \omega_1, \ldots)$ действительных чисел; пусть $\mathscr{F}$ есть σ-поле, порожденное цилиндрами, т. е. множествами вида $\{\omega: (x_n(\omega), \ldots, x_{n+k-1}(\omega)) \in E\}$, где E есть k-мерное борелевское множество (здесь координатными переменными являются $x_n(\omega) = \omega_n$); и, наконец, пусть P — любая мера, которая сохраняется при сдвиге T, определенном формулой $x_n(T\omega) = x_{n+1}(\omega)$. Мера P определяется посредством конечномерных мер

$$\mu_k(E) = P\{\omega: (x_n(\omega), \ldots, x_{n+k-1}(\omega)) \in E\},$$

индуцирующих ее. Сдвиг T соответствует стохастическому процессу $\{\ldots, x_{-1}, x_0, x_1, \ldots\}$ с действительными значениями.

Если каждая мера μ_k является k-кратным произведением μ_1, то $\{x_n\}$ — процесс с независимыми значениями. Тогда в силу теоремы 1.2 преобразование T обладает свойством перемешивания и потому эргодично. Положив в эргодической теореме $f(\omega) = x_0(\omega)$, видим, что при $E\{|x_0|\} < \infty$

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} x_k(\omega) = E\{x_0\} \quad \text{п. в.}$$

Таким образом, эргодическая теорема содержит усиленный закон больших чисел для независимых одинаково распределенных случайных величин с конечным первым моментом.

Пример 1.7 обладает такой общностью, что выполнение требований эргодической теоремы для каждого преобразования такого вида означает выполнение их для любого обратимого преобразования. В самом деле, для того чтобы доказать эргодическую теорему для данной интегрируемой функции $\tilde{f}$ и обратимого сохраняющего меру преобразования $\tilde{T}$ на вероятностном пространстве $(\tilde{\Omega}, \tilde{\mathscr{F}}, \tilde{P})$, возьмем $(\Omega, \mathscr{F})$, как в примере 1.7, и зададим меру P на $\mathscr{F}$ формулой $P = \tilde{P}\varphi^{-1}$, где $\varphi(\tilde{\omega}) = (\ldots, \tilde{f}(\tilde{T}^{-1}\tilde{\omega}), \tilde{f}(\tilde{\omega}), \tilde{f}(\tilde{T}\tilde{\omega}), \ldots)$. Тогда функция $x_0(\omega)$ интегрируема на Ω. Если A есть ω-множество, на котором $\frac{1}{n} \sum_{k=0}^{n-1} x_k(\omega)$ сходится к $\int x_0\, dP$ и $\tilde{A}$ — мно-

[1] Звездочкой отмечены те разделы, которые можно опустить.

жество, на котором $\frac{1}{n}\sum_{k=0}^{n-1}\tilde{f}(\tilde{T}^k\tilde{\omega})$ сходится к $\int \tilde{f}\,d\tilde{P}$, то $\tilde{A} = \varphi^{-1}A$. Если для сдвига T на пространстве Ω справедлива эргодическая теорема, то $P(A) = 1$, так что и $\tilde{P}(\tilde{A}) = 1$. (Если преобразование $\tilde{T}$ необратимо, то аналогичный анализ можно провести с пространством односторонних последовательностей действительных чисел.)

Аналогичное рассуждение показывает, что из эргодической теоремы для сдвига примера 1.2 вытекает, что для любого обратимого эргодического преобразования T каждое множество того пространства, на котором это преобразование определено, является его респектом.

Пример 1.7 можно использовать как образец для сравнения иным образом. Пусть $\{\ldots, \tilde{x}_{-1}, \tilde{x}_0, \tilde{x}_1, \ldots\}$ — стохастический стационарный процесс, определенный на пространстве $(\tilde{\Omega}, \tilde{\mathscr{F}}, \tilde{P})$, т. е. $\tilde{P}\{\tilde{\omega}: (\tilde{x}_n(\tilde{\omega}), \ldots, \tilde{x}_{n+k-1}(\tilde{\omega})) \in E\}$ не зависит от n. Здесь нет необходимости считать $\tilde{\Omega}$ произведением пространств, а $\tilde{x}_n$ — координатными переменными. Определим отображение ψ нашего пространства $\tilde{\Omega}$ в пространство Ω примера 1.7 формулой $\psi(\tilde{\omega}) = \{\ldots, \tilde{x}_{-1}(\tilde{\omega}), \tilde{x}_0(\tilde{\omega}), \tilde{x}_1(\tilde{\omega}), \ldots\}$, и пусть $P = \tilde{P}\psi^{-1}$. Тогда сдвиг T на пространстве Ω сохраняет меру P в силу предположения о стационарности процесса $\{\tilde{x}_n\}$. Все определения и вопросы, касающиеся $\{\tilde{x}_n\}$, могут быть сформулированы в терминах координатных переменных x_n в пространстве Ω и меры P. Например, процесс $\{\tilde{x}_n\}$ называют эргодическим или перемешивающим в зависимости от того, обладает ли соответствующим свойством сдвиг T; если $\{\tilde{x}_n\}$ — эргодический в этом смысле процесс и $\int |\tilde{x}_0|\,d\tilde{P} < \infty$, то, применяя эргодическую теорему к сдвигу T, получаем, что $\frac{1}{n}\sum_{k=0}^{n-1}\tilde{x}_k(\tilde{\omega}) \to \int \tilde{x}_0\,d\tilde{P}$ с вероятностью единица.

З а м е ч а н и е. Наше введение в эргодическую теорию игнорирует факт происхождения ее из статистической механики (см. Кац [1]).

Общее изложение эргодической теории можно найти в работах Халмоша [3], Хопфа [1] и Якобса [1, 2] [1]).

---

[1]) См. также Рохлин [6]. Современное состояние эргодической теории отражено в серии статей, опубликованных в *УМН*, **21**, вып. 5 (1967). — *Прим. ред.*

2. ДОКАЗАТЕЛЬСТВО ЭРГОДИЧЕСКОЙ ТЕОРЕМЫ

Эргодическая теорема заслуживает более чем одного доказательства. В этом параграфе мы сначала сведем эргодическую теорему к так называемой максимальной эргодической теореме (теорема 2.4), используя функциональные пространства L^1 и L^2. Затем докажем максимальную эргодическую теорему безотносительно к L^1 и L^2. И, наконец, дадим еще одно доказательство, опирающееся на максимальную эргодическую теорему и не использующее пространств L^1 и L^2.

Первое доказательство [1])

Сохраняющее меру преобразование T на вероятностном пространстве $(\Omega, \mathcal{F}, P)$ естественным образом порождает оператор $\widehat{T}$ в гильбертовом пространстве L^2 действительнозначных, интегрируемых с квадратом функций, определенных на Ω. Пусть $f(\omega)$ — любая действительная функция на пространстве Ω. Будем обозначать символом $\widehat{T}f$ функцию, которая принимает в точке ω значение $f(T\omega)$. Если f измерима (все рассматриваемые ниже функции будем предполагать измеримыми), то и функция $\widehat{T}f$ измерима, так как T измеримо; если f принадлежит L^2, то и $\widehat{T}f$ принадлежит L^2, ибо преобразование T сохраняет меру. Итак, оператор $\widehat{T}$ отображает пространство L^2 на себя; линейность его очевидна. С помощью замены переменной имеем [2])

$$(\widehat{T}f, \widehat{T}g) = \int \widehat{T}f \cdot \widehat{T}g \, dP = \int f(T\omega) g(T\omega) P(d\omega) = \\ = \int f(\omega) g(\omega) P(d\omega) = (f, g).$$

Скалярные произведения и расстояния сохраняются: $\widehat{T}$ — изометрический оператор [3]).

Пространство интегрируемых с квадратом функций на пространстве пяти точек $\{a, b, c, d, e\}$ из примеров 1.3 и 1.4

---

[1]) Можно, минуя первое доказательство эргодической теоремы, сразу заняться теоремой 2.4. Но доказательство это поучительно, хотя и ведет нас окольным путем.

[2]) Символом (f, g) мы обозначаем скалярное произведение, символом $\|f\|_2 = \sqrt{(f, f)}$ — норму в пространстве L^2.

[3]) Изучение изометрического оператора $\widehat{T}$, лишь самые простые свойства которого здесь используются, может дать совсем немного сведений о порождающем его преобразовании T, сохраняющем меру. Дальнейшая информация относительно $\widehat{T}$ содержится в конце § 5.

можно отождествить с пространством R^5 5-мерных векторов (столбцов). При таком отождествлении изометрические операторы $\widehat{T}$, порожденные преобразованиями этих примеров, задаются матрицами перестановок

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \quad \text{и} \quad \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

соответствующими $(a, b, c)(d, e)$ и (a, b, c, d, e). Поворот окружности (пример 1.5) порождает изометрический оператор, переводящий функцию с рядом Фурье $\sum_n a_n e_n(\omega)$ в функцию с рядом Фурье $\sum_n c^n a_n e_n(\omega)$.

Эргодическая теорема утверждает, что если функция f интегрируема, то

$$\frac{1}{n}\sum_{k=0}^{n-1} f(T^k\omega) \to \hat{f}(\omega) \quad \text{п. в.}$$

для соответствующей функции $\hat{f}$. Мы докажем сначала менее сильный результат, известный под названием эргодической теоремы[1]) в L^2.

Обозначим символом A_n оператор усреднения

$$A_n f = \frac{1}{n}\sum_{k=0}^{n-1} \widehat{T}^k f.$$

Заметим, что

$$\| A_n f \|_2 \leqslant \| f \|_2. \tag{2.1}$$

Теорема 2.1. *Для любой функции f из L^2 существует инвариантная функция $\hat{f}$ из L^2, такая, что $A_n f \to \hat{f}$ в смысле L^2, т. е. $\| A_n f - \hat{f} \|_2 \to 0$.*

Доказательство. Если $A_n f$ вообще сходится к функции из L^2, то эта предельная функция, разумеется, инвариантна. Пусть E^2 — система функций f из L^2, для которых последовательность $\{A_n f\}$ сходится. Мы покажем, что E^2 замкнуто и содержит подмножество, порождающее L^2, а следовательно, совпадает с L^2.

[1]) Для того чтобы отличить саму эргодическую теорему (теорема 1.3) от прочих, ее называют иногда индивидуальной эргодической теоремой.

Пусть $f_k \in E^2$ и $\| f_k - f \|_2 \to 0$. Тогда и $f \in E^2$. Чтобы доказать это, достаточно проверить, что средние арифметические $A_n f$ образуют фундаментальную последовательность. Если $\| f_k - f \|_2 < \varepsilon$, то в силу (2.1)

$$\| A_m f - A_n f \|_2 \leqslant \| A_m f - A_m f_k \|_2 + \| A_m f_k - A_n f_k \|_2 + \\ + \| A_n f_k - A_n f \|_2 < \| A_m f_k - A_n f_k \|_2 + 2\varepsilon.$$

Последовательность $\{A_n f\}$, действительно, фундаментальна, так как $\| A_m f_k - A_n f_k \|_2 \to 0$ при $m, n \to \infty$. Следовательно, система E^2 замкнута.

Если $f = \hat{T} f$, то, очевидно, $f \in E^2 (\bar{f} = f)$. Если f имеет вид $f = g - \hat{T} g$, то $A_n f = (g - \hat{T}^n g)/n$, так что $\| A_n f \|_2 \leqslant 2 \| g \|_2 / n \to 0$ и $f \in E^2 (\bar{f} = 0)$. Таким образом, E^2 содержит класс E_0^2, состоящий из всех инвариантных функций и всех функций вида $g - \hat{T} g$.

Остается еще показать, что E_0^2 порождает L^2, т. е. единственная ортогональная к E_0^2 функция есть 0. Для этого достаточно показать, что любая функция h, ортогональная всякой функции вида $g - \hat{T} g$, удовлетворяет требованию $h = \hat{T} h$ (тогда если она ортогональна и всякой инвариантной функции, то она ортогональна самой себе и, следовательно, равна 0). Благодаря свойству $(f_1, \hat{T} f_2) = (\hat{T}^* f_1, f_2)$ сопряженного оператора $\hat{T}^*$ имеет место тождество $(h, g - \hat{T} g) = (h - \hat{T}^* h, g)$. Поэтому если $(h, g - \hat{T} g) = 0$ для всех g, то $h = \hat{T}^* h$. Далее[1]), $\| h - \hat{T} h \|_2^2 = \| h \|_2^2 - (h, \hat{T} h) - (\hat{T} h, h) + \| \hat{T} h \|_2^2$. Очевидно, что $\| \hat{T} h \|_2^2 = \| h \|_2^2$, и если $h = \hat{T}^* h$, то $(h, \hat{T} h) = (\hat{T}^* h, h) = \| h \|_2^2$ и аналогично $(\hat{T} h, h) = \| h \|_2^2$, так что $\| h - \hat{T} h \|_2^2 = 0$, или $h = \hat{T} h$. Итак, $h = \hat{T} h$, если функция h ортогональна каждой функции вида $g - \hat{T} g$, что завершает доказательство.

Заметим, что оператор, переводящий f в $\bar{f}$, является ортогональным проектором на подпространство инвариантных функций.

Следующий шаг состоит в доказательстве теоремы, аналогичной теореме 2.1, для пространства L^1. Так как $P(\Omega) = 1$, то L^2 является подпространством в L^1. Если функция $f(\omega)$ интегрируема, то интегрируема и $f(T\omega)$ и, следовательно,

---

[1]) Если изометрический оператор $\hat{T}$ обратим, то он унитарен, так что из $h = \hat{T}^* h = \hat{T}^{-1} h$ следует выполнение требования $h = \hat{T} h$. Однако $\hat{T}$ может быть необратимым, что обосновывает необходимость дальнейших рассуждений.

оператор $\widehat{T}$ определен на пространстве L^1 интегрируемых функций. Теперь мы будем рассматривать $\widehat{T}$ как оператор в L^1, отображающий подпространство L^2 в себя. Если $\| f \|_1 = \int | f | dP$ обозначает норму в L^1, то $\| \widehat{T} f \|_1 = \| f \|_1$ и (мы используем символ A_n, как и выше, для обозначения среднего $\frac{1}{n} \sum_{k=0}^{n-1} \widehat{T}^k$, но теперь в L^1) $\| A_n f \|_1 \leqslant \| f \|_1$.

Теорема 2.2. *Для любой функции f из L^1 существует инвариантная функция $\hat{f}$ из L^1, такая, что $A_n f \to \hat{f}$ в смысле L^1, т. е. $\| A_n f - \hat{f} \|_1 \to 0$.*

Доказательство. Обозначим через E^1 множество функций f из L^1, для которых утверждение теоремы справедливо. Те же рассуждения, с помощью которых мы показали замкнутость системы E^2 в пространстве L^2, доказывают и замкнутость E^1 в L^1 — нужно только везде заменить норму в L^2 на норму в L^1. Если f принадлежит L^2, то по предыдущей теореме существует некоторая функция $\hat{f}$ из L^2, такая, что $\| A_n f - \hat{f} \|_2 \to 0$, и эта функция принадлежит, разумеется, и L^1. В силу неравенства Гельдера имеем $\| A_n f - \hat{f} \|_1 \leqslant \| A_n f - \hat{f} \|_2 \to 0$, так что $f \in E^1$. Итак, $L^2 \subset E^1$. Любая функция f из L^1 является предельной в смысле L^1 для функций из L^2 (ограниченных), так что замыкание L^2 в L^1 совпадает с самим L^1. Так как $L^2 \subset E^1$ и E^1 замкнуто, то $E^1 = L^1$.

Очевидно, что $\int A_n f \, dP = \int f \, dP$, и так как $A_n f \to \hat{f}$ в смысле L^1, то

$$\int \hat{f} \, dP = \int f \, dP. \tag{2.2}$$

Попытаемся построить доказательство индивидуальной эргодической теоремы, следуя линии двух предыдущих доказательств. Пусть G — множество элементов f из L^1, для которых существует функция $\hat{f}$ такая, что

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k \omega) = \hat{f}(\omega) \text{ п. в.}$$

Если такая функция $\hat{f}$ вообще существует, она должна совпадать с функцией $\hat{f}$ из теоремы 2.2 и, следовательно, должна быть инвариантным элементом пространства L^1, удовлетворяющим соотношению (2.2). Мы докажем индиви-

дуальную эргодическую теорему, показав, что G (a) замкнуто в L^1 и (b) содержит подмножество, порождающее L^1.

Предположим, что утверждение (a) справедливо, и будем доказывать (b). Если $f=\hat{T}f$, то $f\in G$ $(\tilde{f}=f)$. Если $f=g-\hat{T}g$, где $g(\omega)$ — ограниченная (скажем, константой K) функция, то

$$|A_nf(\omega)|=|g(\omega)-g(\hat{T}^n\omega)|/n\leqslant 2K/n\to 0$$

для каждого ω, так что $f\in G$ $(\tilde{f}=0)$. Если $f=g-\hat{T}g$, где $g\in L^1$, выберем ограниченные функции g_k такие, что $\|g_k-g\|_1\to 0$. В силу непрерывности оператора $\hat{T}$ функция f является предельной в смысле L^1 для функций $g_k-\hat{T}g_k$ и, значит (если выполняется (a)), принадлежит G. Следовательно, G содержит класс E_0^2, введенный в доказательстве теоремы 2.1, состоящий из всех инвариантных элементов L^2 и всех элементов вида $g-\hat{T}g$, где g принадлежит L^2. Было показано, что замыкание класса E_0^2 в L^2 есть само L^2. В силу неравенства Гельдера любой предельный в смысле L^2 элемент E_0^2 является также предельным в смысле L^1; таким образом, множество G, по предположению замкнутое в L^1, содержит L^2 и потому совпадает с L^1.

Остается показать, что G замкнуто в смысле L^1. Будем считать установленным следующий результат.

Теорема 2.3. *Если* $f\in L^1$ *и* $\lambda>0$, *то*

$$P\left\{\omega:\ \sup_{n\geqslant 1}|A_nf(\omega)|>\lambda\right\}\leqslant\frac{1}{\lambda}\|f\|_1.$$

Предположим, что $f_k\in G$ и $\|f_k-f\|_1\to 0$. Для каждого ω

$$|A_mf(\omega)-A_nf(\omega)|\leqslant|A_mf_k(\omega)-A_nf_k(\omega)|+ \\ +2\sup_{v\geqslant 1}|A_v(f(\omega)-f_k(\omega))|. \qquad (2.3)$$

Так как $f_k\in G$, то последовательность $\{A_nf_k(\omega)\}$ фундаментальна почти всюду для каждого k, так что первый член в правой части неравенства (2.3) стремится к 0, когда $m,\ n\to\infty$. Тогда в силу теоремы 2.3

$$\overline{\lim_{m,\,n\to\infty}}|A_mf(\omega)-A_nf(\omega)|=\lim_{k\to\infty}\sup_{m,\,n\geqslant k}|A_mf(\omega)-A_nf(\omega)| \qquad (2.4)$$

превышает λ с вероятностью, не большей $2\|f-f_k\|_1/\lambda$. Устремляя $k\to\infty$ и $\lambda\to 0$, видим, что значение (2.4) поло-

жительно с вероятностью 0: последовательность $\{A_n f(\omega)\}$ фундаментальна почти всюду. Итак, $A_n f(\omega)$ имеет почти всюду предел, который в силу теоремы 2.2 должен быть инвариантным элементом пространства L^1.

Первое доказательство эргодической теоремы закончено. Заметим, что теорема 2.3 играет ту же роль в доказательстве замкнутости G, что и элементарное неравенство (2.1) в доказательствах замкнутости многообразий E^2 и E^1 (теоремы 2.1 и 2.2). Так как в индивидуальной эргодической теореме мы имеем дело со сходимостью почти всюду, то мы не можем использовать неравенства, которые существенно слабее неравенства теоремы 2.3. Теорема 2.3 следует из теоремы 2.4 (заменой f на $|f|$).

Максимальная эргодическая теорема

Итак, все сводится к следующему результату, известному под названием максимальной эргодической теоремы.

Теорема 2.4. *Пусть*

$$N = \left\{ \omega:\ \sup_{n \geqslant 1} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k \omega) > \lambda \right\}. \tag{2.5}$$

Если f интегрируема, то

$$\lambda P(N) \leqslant \int_N f\, dP. \tag{2.6}$$

Заметим, что можно не требовать положительности λ.

Доказательство. Пусть

$$G = \left\{ \omega:\ \sup_{n \geqslant 1} \sum_{k=0}^{n-1} f(T^k \omega) > 0 \right\}.$$

Достаточно показать, что

$$\int_G f\, dP \geqslant 0;$$

тогда (2.6) получается заменой f на $f - \lambda$.

Если

$$G_k = \left\{ \omega:\ \max_{1 \leqslant u \leqslant k} \sum_{i=0}^{u-1} f(T^i \omega) > 0 \right\},$$

то $\int\limits_{G_k} f\,dP \to \int\limits_{G} f\,dP$ и, следовательно, в силу теоремы об арифметических средних $\frac{1}{n}\sum\limits_{k=1}^{n}\int\limits_{G_k} f\,dP \to \int\limits_{G} f\,dP$. Поэтому достаточно доказать, что

$$\sum_{k=1}^{n}\int\limits_{G_k} f\,dP \geqslant 0, \quad n=1,\ 2,\ \ldots. \tag{2.7}$$

При $n=1$ это неравенство тривиально, оно даже не содержит T. Поучительно детальное рассмотрение случая $n=2$. В этом случае неравенство (2.7) записывается как

$$\int\limits_{G_1} f(\omega)\,dP + \int\limits_{G_2} f(\omega)\,dP \geqslant 0, \tag{2.8}$$

где $G_1=\{f(\omega)>0\}$ и $G_2=\{f(\omega)>0$ или $f(\omega)+f(T\omega)>0\}$. Если $G_1'=\{f(T\omega)>0\}$, то с помощью замены переменной находим $\int\limits_{G_1} f(\omega)\,dP = \int\limits_{G_1'} f(T\omega)\,dP$ (мы используем здесь сохранение меры преобразованием T), так что неравенство (2.8) эквивалентно неравенству

$$\int\limits_{G_2} f(\omega)\,dP + \int\limits_{G_1'} f(T\omega)\,dP \geqslant 0 \tag{2.9}$$

и, следовательно, неравенству

$$\int\limits_{G_2\cap G_1'} [f(\omega)+f(T\omega)]\,dP + \int\limits_{G_2-G_1'} f(\omega)\,dP + \int\limits_{G_1'-G_2} f(T\omega)\,dP \geqslant 0.$$

Неравенство это выполняется, так как каждый из трех содержащихся в нем интегралов неотрицателен, поскольку подинтегральная функция в каждом из них положительна в области интегрирования. Можно рассуждать несколько иным образом, замечая сначала, что (2.9) эквивалентно неравенству

$$\int \left[f(\omega)\,I_{G_2}(\omega) + f(T\omega)\,I_{G_1'}(\omega)\right] dP \geqslant 0,$$

а затем, что подинтегральная функция в этом выражении неотрицательна для каждого ω. Такое рассуждение применимо и в случае $n>2$.

Фиксируем n, и пусть $H_u = T^{-u} G_{n-u}$. С помощью замены переменной имеем

$$\sum_{k=1}^{n} \int_{G_k} f(\omega)\, dP = \sum_{u=0}^{n-1} \int_{G_{n-u}} f(\omega)\, dP = \sum_{u=0}^{n-1} \int_{H_u} f(T^u \omega)\, dP =$$

$$= \int \left\{ \sum_{n=0}^{n-1} f(T^u \omega) I_{H_u}(\omega) \right\} dP.$$

(Этот шаг соответствует сделанному выше переходу от G_1 к G_1'; здесь мы используем предположение о сохранении меры преобразованием T.) Неравенство (2.7) будет доказано если мы покажем, что подинтегральное выражение

$$\sum_{u=0}^{n-1} f(T^u \omega) I_{H_u}(\omega) \tag{2.10}$$

неотрицательно в каждой точке ω. Точка ω' принадлежит множеству G_k в том и только том случае, если хотя бы одна из сумм

$$f(\omega'),\ f(\omega') + f(T\omega'),\ \ldots,\ f(\omega') + \ldots + f(T^{k-1}\omega')$$

положительна; следовательно, ω' принадлежит G_{n-u} в том и только том случае, если хотя бы одна из сумм

$$f(\omega'),\ f(\omega') + f(T\omega'),\ \ldots,\ T(\omega') + \ldots + f\left(T^{n-u-1}\omega'\right)$$

положительна. Следовательно, $\omega \in H_u = T^{-u} G_{n-u}$ или $T^u \omega \in \in G_{n-u}$ в том и только том случае, если хотя бы одна из сумм

$$f\left(T^u \omega\right),\ f\left(T^u \omega\right) + f\left(T^{u+1}\omega\right),\ \ldots,\ f\left(T^u \omega\right) + \ldots + f\left(T^{n-1}\omega\right) \tag{2.11}$$

положительна. Таким образом, (2.10) — сумма $f(T^u \omega)$ по тем u, для которых хотя бы одна из сумм (2.11) положительна.

Зафиксируем ω и положим $c_u = f(T^u \omega)$. Доказательство максимальной эргодической теоремы завершается следующей комбинаторной леммой.

Лемма. *Назовем член c_u конечной последовательности чисел $c_0, c_1, \ldots, c_{n-1}$ лидером, если хотя бы одна из сумм*

$$c_u,\ c_u + c_{u+1},\ \ldots,\ c_u + \ldots + c_{n-1}$$

положительна. Тогда сумма[1]) *лидеров неотрицательна.*

---

[1]) Пустую сумму условимся считать нулем.

Доказательство получается индукций по числу n элементов последовательности. Для $n=1$ результат тривиален. Предположим, что он справедлив для натуральных чисел, меньших n. Если c_0 не лидер, то все лидеры данной последовательности являются одновременно лидерами укороченной последовательности $c_1, \ldots, c_{n-1}$ и их сумма неотрицательна в силу предположения индукции. Пусть теперь c_0 — лидер. Если k — наименьшее натуральное число, для которого сумма $c_0 + \ldots + c_k$ положительна, то $c_1, \ldots, c_k$ — тоже лидеры, ибо если бы один из них, скажем c_i, не являлся лидером, то сумма $c_i + \ldots + c_k$ не была бы положительной и, значит, $c_0 + \ldots + c_{i-1} > 0$, что противоречит выбору k. Итак, сумма лидеров $c_0, c_1, \ldots, c_k$ положительна; сумма остальных лидеров (если такие существуют) неотрицательна, так как они лидируют в последовательности, число членов которой меньше n. Доказательство завершено.

Второе доказательство

Эргодическую теорему можно непосредственно вывести из максимальной эргодической теоремы. Заметим, что если множество A инвариантно, то

$$\lambda P(A \cap N) \leqslant \int\limits_{A \cap N} f\, dP \tag{2.12}$$

(где N определено выражением (2.5) и f интегрируема), что легко получается заменой f на $f \cdot I_A$. Используя это замечание, покажем, что с точностью до множества меры 0 последовательность средних арифметических

$$\left\{ \frac{1}{n} \sum_{k=0}^{n-1} f(T^k\omega) \right\}$$

сходится (быть может, к $+\infty$ или $-\infty$).

Положим для $a < b$

$$A_{a,b} = \left\{ \varliminf_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k\omega) < a < b < \varlimsup_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k\omega) \right\}.$$

Очевидно, что $A_{a,b}$ инвариантно и

$$A_{a,b} = A_{a,b} \cap \left\{ \sup_{n \geqslant 1} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k\omega) \geqslant b \right\}.$$

Если в (2.12) положить $\lambda = b$, имеем

$$bP(A_{a,b}) \leqslant \int\limits_{A_{a,b}} f\,dP.$$

Если вместо f, a, b взять $-f$, $-a$, $-b$, получим

$$\int\limits_{A_{a,b}} f\,dP \leqslant aP(A_{a,b}).$$

Таким образом, $bP(A_{a,b}) \leqslant aP(A_{a,b})$, что возможно, только если $P(A_{a,b}) = 0$, поскольку $a < b$. Если A — объединение $A_{a,b}$ по всем парам рациональных a и b $(a < b)$, то $P(A) = 0$.

Вне A значения верхнего и нижнего пределов последовательности средних совпадают. Итак,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k\omega) = \bar{f}(\omega) \quad \text{п. в.},$$

где $\bar{f}(\omega)$ конечно или равно $+\infty$ или $-\infty$. С помощью замены переменной получаем

$$\int \frac{1}{n} \sum_{k=0}^{n-1} |f(T^k\omega)|\,dP = \int |f|\,dP$$

и, применяя лемму Фату, имеем

$$\int |\bar{f}|\,dP \leqslant \varliminf_{n\to\infty} \int \left| \frac{1}{n} \sum_{k=0}^{n-1} f(T^k\omega) \right| dP \leqslant \int |f|\,dP.$$

Таким образом, функция $\bar{f}$ интегрируема; в частности, она почти всюду имеет конечное значение.

Так как предельная функция $\bar{f}$ инвариантна, остается только показать, что интеграл от нее совпадает с интегралом от f. Обозначим $a_n(\omega) = \frac{1}{n} \sum\limits_{k=0}^{n-1} f(T^k\omega)$. Так, как $a_n(\omega) \to \bar{f}(\omega)$ почти всюду и $\int a_n\,dP = \int f\,dP$, то, производя формальные действия, получаем

$$\int \bar{f}\,dP = \int \lim_{n\to\infty} a_n\,dP = \lim_{n\to\infty} \int a_n\,dP = \int f\,dP. \qquad (2.13)$$

Покажем, что такая перестановка предела и интеграла законна[1]). Если $N_\lambda = \left\{\omega: \sup\limits_{n \geqslant 1} |a_n(\omega)| > \lambda\right\}$, то

$$\int |a_n - \bar{f}|\,dP \leqslant \int\limits_{N_\lambda^c} |a_n - \bar{f}|\,dP + \int\limits_{N_\lambda} |a_n|\,dP + \int\limits_{N_\lambda} |\bar{f}|\,dP.$$

При фиксированном положительном λ первый член суммы в правой части неравенства стремится к 0, когда $n \to \infty$, в силу теоремы о мажорированной сходимости. Поэтому достаточно показать, что второй и третий члены суммы можно сделать малыми равномерно по n, выбрав большое λ. Относительно третьего члена это утверждение справедливо, потому что в силу максимальной эргодической теоремы $P(N_\lambda) \leqslant E\{|f|\}/\lambda \to 0$, когда $n \to \infty$. Что касается второго члена, то мы имеем

$$\int\limits_{N_\lambda} |a_n|\,dP \leqslant \frac{1}{n} \sum_{k=0}^{n-1} \int\limits_{N_\lambda} |f(T^k\omega)|\,P(d\omega)$$

и

$$\int\limits_{N_\lambda} |f(T^k\omega)|\,P(d\omega) \leqslant \int\limits_{\{|f(T^k\omega)| > \alpha\}} |f(T^k\omega)|\,P(d\omega) + \alpha P(N_\lambda),$$

так что с помощью замены переменной получаем

$$\int\limits_{N_\lambda} |a_n|\,dP \leqslant \int\limits_{\{|f(\omega)| > \lambda\}} |f(\omega)|\,P(d\omega) + \alpha P(N_\lambda).$$

Мы можем сделать правую часть этого неравенства малой, выбрав достаточно большие α и λ.

Итак, $\int |a_n - \bar{f}|\,dP \to 0$, что делает законной перестановку предела и интеграла в (2.13)[2]) и полностью завершает доказательство эргодической теоремы.

Замечание. Эргодическая теорема в L^2 была впервые доказана Нейманом [1], индивидуальная эргодическая теорема — Биркгофом [1]. Приведенное нами доказательство максимальной эргодической теоремы принадлежит Риссу [1].

---

[1]) Это очевидно, если f ограничена. Следовательно, уже можно заключить, что если T эргодично, то траектории воспроизводят Ω в том смысле, в котором речь об этом шла в § 1.

[2]) И, между прочим, еще раз доказывает теорему 2.2. Незначительные изменения в рассуждениях приводят ко второму доказательству теоремы 2.1.

3. ДАЛЬНЕЙШИЕ ПРИМЕРЫ

В этом параграфе мы собрали несколько примеров, которые понадобятся нам позже для иллюстрации различных результатов.

Сдвиги

Пример 3.1. Рассмотрим специальный случай общего сдвига (пример 1.2). Пусть $\Pi=(p_{ij})$ — стохастическая $(r\times r)$-матрица, строки и столбцы которой соответствуют элементам пространства состояний ρ. Обозначим через $p=(p_i)$ такой вектор (строку) вероятностей, что $p\Pi=p$. Мы не делаем никаких специальных предположений относительно матрицы Π; она может быть или не быть неприводимой, непериодической и т. д. [1]).

Функции

$$p_k(i_1, \ldots, i_k)=p_{i_1}p_{i_1i_2}\cdots p_{i_{k-1}i_k}$$

удовлетворяют условиям (1.3), первое из которых очевидно, второе следует из стохастичности матрицы Π, третье — из того, что p — вектор вероятностей. Условие стационарности (1.4) также выполняется, поскольку $p\Pi=p$. Поэтому существует единственная мера P, сохраняющаяся при сдвиге T, такая, что

$$P\left\{x_n(\omega)=i_1, \ldots, x_{n+k-1}(\omega)=i_k\right\}=p_{i_1}p_{i_1i_2}\cdots p_{i_{k-1}i_k}.$$

При так определенной мере мы будем называть T *сдвигом Маркова* [2]). Сдвиг Бернулли является частным случаем сдвига Маркова при $p_{ij}=p_j$. Подобным же образом можно определить сдвиг Маркова более высокого порядка.

В силу теоремы 1.2 общий сдвиг T является перемешивающим в том и только том случае, если соотношение $\lim\limits_n P(A\cap T^{-n}B)=P(A)P(B)$ выполняется для всех цилиндров A и B. Так как каждый цилиндр представляет собой конечное объединение непересекающихся тонких цилиндров, то достаточно, чтобы это соотношение выполнялось для

---

[1]) В терминологии мы следуем Феллеру [1].

[2]) В отечественной литературе принят термин *автоморфизм Маркова*. — *Прим. ред.*

тонких цилиндров A и B: сдвиг T является перемешивающим в том и только том случае, если для всех i и j

$$\lim_{n\to\infty} P\{x_1 = i_1, \ldots, x_u = i_u;\ x_{n+1} = j_1, \ldots, x_{n+v} = j_v\} = \\ = P\{x_1 = i_1, \ldots, x_u = i_u\} P\{x_1 = j_1, \ldots, x_v = j_v\}. \quad (3.1)$$

Аналогичные рассуждения, опирающиеся на теорему 1.4, показывают, что сдвиг T эргодичен в том и только том случае, если

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} P\{x_1 = i_1, \ldots, x_u = i_u;\ x_{k+1} = j_1, \ldots, x_{k+v} = j_v\} = \\ = P\{x_1 = i_1, \ldots, x_u = i_u\} P\{x_1 = j_1, \ldots, x_v = j_v\}. \quad (3.2)$$

Согласно (1.9), предел

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} P\{x_1 = i_1, \ldots, x_u = i_u;\ x_{k+1} = j_1, \ldots, x_{k+v} = j_v\} \quad (3.3)$$

существует, даже если T не эргодичен.

Вернемся к сдвигу Маркова. Будем предполагать теперь, что $p_i > 0$ для всех i, принадлежащих ρ (любое состояние i с $p_i = 0$ никогда не может наступить и потому не имеет отношения к процессу). Изучим условия, при которых сдвиг Маркова эргодичен.

Полагая в соотношении (3.3) $u = v = 1$, видим, что предел

$$q_{ij} = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} p_{ij}^{(k)}$$

всегда существует[1]). Если T эргодичен, то $q_{ij} = p_j$ в силу (3.2). С другой стороны, если $p_{ij} = p_j$, то сдвиг T эргодичен, так как вероятность в соотношении (3.2) равна

$$p_{i_1} p_{i_1 i_2} \ldots p_{i_{u-1} i_u} p_{i_u j_1}^{(k+1-u)} p_{j_1 j_2} \ldots p_{j_{v-1} j_v}, \qquad k \geqslant u.$$

Докажем эквивалентность следующих четырех утверждений:

(a) *сдвиг T эргодичен*;

(b) *q_{ij} не зависит от i*;

---

[1]) Здесь $p_{ij}^{(k)}$ означает вероятность перехода из состояния i в состояние j за k шагов, причем $p_{ij}^{(0)} = \delta_{ij}$.

(c) *матрица* Π *неприводима*[1]);

(d) $q_{ij} > 0$ *для любой пары* i, j.

Пусть числа q_{ij} образуют матрицу Q:

$$Q = \lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} \Pi^k. \tag{3.4}$$

Нетрудно показать, что Q — стохастическая матрица, для которой выполняются соотношения

$$Q\Pi = \Pi Q = Q, \quad Q^2 = Q. \tag{3.5}$$

Так как $p\Pi = p$, то $p\Pi^k = p$ для всех k, следовательно, в силу (3.4) $pQ = p$. Значит, q_{ij} не зависят от i в том и только том случае, если $q_{ij} = p_j$. Таким образом, установлена эквивалентность утверждений (a) и (b).

Предположим, что матрица Π не является неприводимой; обозначим через ρ_0 собственное замкнутое подмножество в ρ. Пусть $A = \{\omega: x_0(\omega) \in \rho_0\}$, тогда $0 < P(A) < 1$ и $P(A - T^{-1}A) = 0$. Так как A — инвариантное множество (см. замечание, следующее за определением инвариантности в § 1), то сдвиг T не эргодичен. Итак, из (a) вытекает (c).

В силу (3.5) $q_{ij} = \sum_u q_{iu} p_{uj} \geqslant q_{ik} p_{kj}$, следовательно, если $q_{ik} > 0$ и $p_{kj} > 0$, то $q_{ij} > 0$. Поэтому ρ_j (множество таких j, что $q_{ij} > 0$) замкнуто. Если матрица Π неприводима, то ρ_i, будучи непустым, должно совпадать с ρ. Таким образом, из (c) вытекает (d).

Предположим теперь, что все q_{ij} положительны и

$$\sum_j q_{ij} \xi_j = \xi_i, \quad i \in \rho. \tag{3.6}$$

Пусть максимальное из ξ_i равно m. Если $\xi_i < m$ для некоторого i, то $\xi_k = \sum_j q_{kj} \xi_j < \sum_j q_{kj} m = m$ для всех k, что невозможно. Поэтому в силу утверждения (d) (3.6) выполняется в том и только том случае, если все ξ_i равны между собой. Так как $Q^2 = Q$, то каждый столбец матрицы Q

---

[1]) Множество $\rho_0 \subset \rho$ замкнуто, если $\sum_{j \in \rho_0} p_{ij} = 1$ для всех i из ρ_0; матрица Π неприводима, если ρ не содержит собственного замкнутого подмножества. Требованием, эквивалентным требованию неприводимости, является существование для каждой пары i, j такого n, что $p_{ij}^{(n)} > 0$.

является решением системы уравнений (3.6); следовательно, q_{ij} не зависят от i. Итак, (d) влечет (b).

Мы показали, что наши четыре утверждения эквивалентны друг другу и что они эквивалентны требованию единственности решения (разумеется, с точностью до множителя) системы (3.6). Легко видеть, что (b) выполняется в том и только том случае, если система

$$\sum_i \eta_i q_{ij} = \eta_j, \quad j \in \rho, \tag{3.7}$$

имеет только одно решение. В силу (3.4) любое решение системы

$$\sum_j p_{ij}\xi_j = \xi_i, \quad i \in \rho, \tag{3.8}$$

является решением системы (3.6) и любое решение системы

$$\sum_i \eta_i p_{ij} = \eta_j, \quad j \in \rho, \tag{3.9}$$

является решением системы (3.7). Таким образом, мы получили два условия, каждое из которых эквивалентно эргодичности T.

(e) *Система* (3.8) *имеет единственное решение.*

(f) *Система* (3.9) *имеет единственное решение.*

Эквивалентность условий (e) и (f) следует также из сопряженности системы (3.9) системе (3.8).

Мы показали, что сдвиг T эргодичен в том и только том случае, если $q_{ij} = p_j$; доказательство опирается на (3.2). Заменяя в наших рассуждениях (3.2) на (3.1), видим, что T обладает свойством перемешивания в том и только том случае, если

$$\lim_{n\to\infty} p_{ij}^{(n)} = p_j, \quad i, j \in \rho. \tag{3.10}$$

Из теории цепей Маркова известно, что (3.10) выполняется в том и только том случае, если матрица Π неприводима и непериодична[1]). Здесь мы воспользуемся этим фактом без доказательства (доказательство см. в § 11).

Пример 3.2. Преобразование сдвига можно определить на произведении элементов односторонней последовательности экземпляров пространства ρ. Тогда элементом про-

---

[1]) Не существует натурального $m > 1$, такого, что $p_{ii}^{(n)} > 0$ только для n, кратного m. (В литературе по цепям Маркова возвратное ненулевое непериодическое *состояние* называют иногда эргодическим; см. Феллер [1]. Мы будем избегать такого употребления.)

странства Ω является последовательность $\omega = (\omega_1, \omega_2, \ldots)$, где $\omega_n \in \rho$; координатные переменные определяются, как и выше; $T(\omega_1, \omega_2, \ldots) = (\omega_2, \omega_3, \ldots)$, так что $(T\omega)_n = \omega_{n+1}$. Функции p_k, удовлетворяющие условиям (1.3) и (1.4), однозначно определяют меру P, сохраняющуюся при преобразовании T. Это преобразование называется *односторонним сдвигом*. Односторонние сдвиги Бернулли и Маркова являются, очевидно, частными случаями этого общего одностороннего сдвига. Заметим, что сдвиг в примере 1.2 (который мы теперь можем назвать двусторонним сдвигом) обратим, в то время как односторонний сдвиг необратим.

Пример 3.3. Пусть $P'[P'']$ — мера на пространстве $(\Omega, \mathscr{F})$, где определен двусторонний сдвиг, относительно которой координатные переменные x_{2n} с четными индексами независимы, принимают значение i с вероятностью p_i и $x_{2n} = x_{2n+1}$ $[x_{2n} = x_{2n-1}]$ для всех n с вероятностью 1. Сдвиг T не сохраняет ни P', ни P'', но сохраняет их среднее арифметическое $P = (P' + P'')/2$, что следует из соотношений

$$P'\left(T^{-k}B\right) = \begin{cases} P'(B) & \text{для четного } k, \\ P''(B) & \text{для нечетного } k \end{cases}$$

и

$$P''\left(T^{-k}B\right) = \begin{cases} P''(B) & \text{для четного } k, \\ P'(B) & \text{для нечетного } k. \end{cases}$$

Из этих соотношений следует также, что

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \frac{1}{2} \left[P'(A) P'\left(T^{-k}B\right) + P''(A) P''\left(T^{-k}B\right)\right] = P(A) P(B).$$

Если A и B — цилиндры, то A и $T^{-k}B$ независимы при P' и P'' для достаточно больших k. Следовательно, условие (1.8) выполняется. Тогда в силу теоремы 1.4 сдвиг T эргодичен. Однако T не обладает свойством перемешивания (если только одна из вероятностей p_i не равна 1).

Меры на интервале

Обобщим диадическое преобразование из примера 1.6, заменив основание 2 основанием $r \geqslant 2$.

Пример 3.4. Пусть P — мера Лебега на классе $\mathscr{F}$ борелевских подмножеств полуинтервала $\Omega = [0, 1)$, и пусть преобразование T задано формулой $T\omega = r\omega \pmod 1$. Если

$f(\omega)=i$ на $[i/r,\ (i+1)/r)$, $i=0,\ 1,\ \ldots,\ r-1$, то точка ω имеет по основанию r разложение $\sum\limits_{n=1}^{\infty} f\left(T^{n-1}\omega\right)/r^n$. Как и в случае $r=2$, преобразование T эргодическое (и даже перемешивающее). Применение эргодической теоремы показывает, что почти каждое число нормально по r (содержит все знаки разложения в одинаковой пропорции). Обозначим символом $x_n(\omega)$ n-й знак разложения ω по основанию r $\left(\text{т. е. } f\left(T^{n-1}\omega\right)\right)$; тогда $x_1,\ x_2,\ \ldots$ образуют последовательность независимых случайных величин, где $P\{x_n=i\}=1/r$, $i=0,\ 1,\ \ldots,\ r-1$. Будем называть T *r-адическим преобразованием*.

Пример 3.5. Определим Ω, $\mathscr{F}$ и T так же, как в предыдущем примере, но пусть теперь P — любая мера, которая сохраняется при преобразовании T. Общий вид r-адического интервала (диадического при $r=2$) — $[k/r^n,\ (k+1)/r^n)$; он содержит точку ω в том и только том случае, если первые n знаков $x_1(\omega),\ \ldots,\ x_n(\omega)$ ее разложения (по основанию r) имеют заданные значения. Так как конечные объединения непересекающихся диадических интервалов образуют поле, порождающее $\mathscr{F}$, преобразование T сохраняет меру P в том и только том случае, если оно сохраняет меру каждого диадического интервала, или, что эквивалентно, в том и только том случае, если последовательность $\{x_1,\ x_2,\ \ldots\}$ образует процесс, стационарный относительно P.

Этот пример показывает, что результаты эргодической теории и теории вероятностей можно превращать в результаты, относящиеся к единичному интервалу, и обратно. Например, из эргодической теоремы можно получить теорему о нормальных числах. В качестве примера обратной процедуры выведем специальный случай теоремы существования Колмогорова из того, что функция распределения на единичном интервале задает на нем меру.

Пусть P — мера на единичном интервале; мы сейчас не задаемся вопросом, сохраняется ли она при преобразовании $T\omega=r\omega\ (\mathrm{mod}\ 1)$. Пусть p_k — функция, определенная на последовательностях $(i_1,\ \ldots,\ i_k)$ длины k знаков разложения по основанию r. Она задается формулой

$$p_k(i_1,\ \ldots,\ i_k)=P\{x_1(\omega)=i_1,\ \ldots,\ x_k(\omega)=i_k\}. \tag{3.11}$$

Тогда

$$\begin{cases} p_k(i_1,\ \ldots,\ i_k)\geqslant 0, \\ \sum\limits_i p_{k+1}(i_1,\ \ldots,\ i_k,\ i)=p_k(i_1,\ \ldots,\ i_k), \\ \sum\limits_i p_1(i)=1 \end{cases} \tag{3.12}$$

(эти соотношения формально тождественны условиям (1.3)). С другой стороны, пусть нам даны функции p_k, удовлетворяющие условиям (3.12). Нужно построить на $\mathscr{F}$ вероятностную меру P, удовлетворяющую соотношению (3.11).

Определим функцию F r-адического рационального аргумента $u/r^k \in [0, 1]$ формулой

$$F\left(\frac{u}{r^k}\right) = \sum p_k(i_1, \ldots, i_k), \tag{3.13}$$

где суммирование ведется по тем наборам $i_1, \ldots, i_k$, для которых $\sum_{l=1}^{k} \frac{i_l}{r^l} < \frac{u}{r^k}$. Значения функции в точках u/r^k и ru/r^{k+1} совпадают в силу второго из условий (3.12). Первое и третье из этих условий показывают, что F — неубывающая функция и $F(1) = 1$. Так как сумма, соответствующая точке 0, пуста, то $F(0) = 0$. Сделаем еще одно предположение, состоящее в том, что для любой последовательности знаков $i_1, \ldots, i_k$ разложения по основанию r выполняется соотношение

$$\lim_{v \to \infty} p_{k+v}(i_1, \ldots, i_k, \underbrace{r-1, \ldots, r-1}_{v}) = 0. \tag{3.14}$$

(Если имеет место (3.11), то это последнее условие должно выполняться, ибо не существует точки ω, разложение которой оканчивалось бы на $r-1, r-1, \ldots$.) Из (3.14) следует, что F — непрерывная слева функция r-адического рационального аргумента и, следовательно, может быть продолжена до функции распределения на $[0, 1]$, т. е. до непрерывной слева[1]) неубывающей функции, такой, что $F(0) = 0$ и $F(1) = 1$. Любой функции распределения F на $[0, 1]$ отвечает единственная вероятностная мера P на $\mathscr{F}$, такая, что

$$P[0, x) = F(x), \quad 0 \leqslant x \leqslant 1. \tag{3.15}$$

Отсюда и из определения F немедленно следует (3.11).

Если P и p_k связаны соотношениями (3.11), то преобразование T сохраняет меру P в том и только том случае, если

$$p_k(i_1, \ldots, i_k) = \sum_i p_{k+1}(i, i_1, \ldots, i_k). \tag{3.16}$$

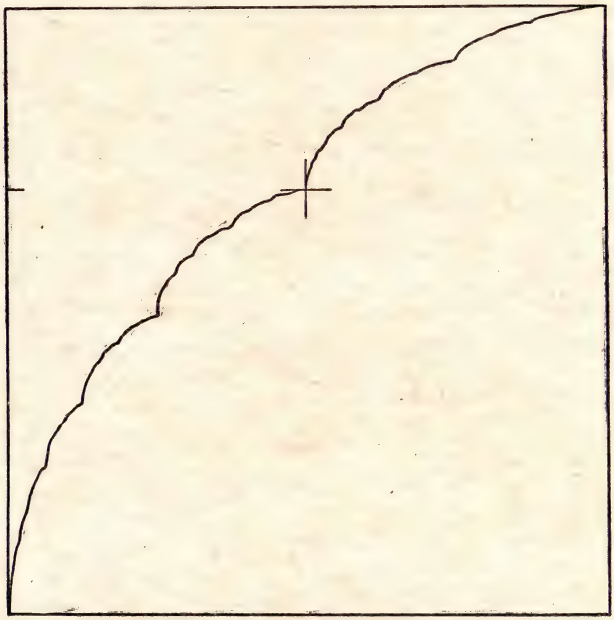
---

[1]) Непрерывность слева является здесь условностью, вызванной тем, что мы имеем дело с интервалами, замкнутыми слева, которые естественно связаны с разложениями, конечными в рациональном случае.

Если $p_0, \ldots, p_{r-1}$ — неотрицательные числа, меньшие 1, в сумме дающие 1, и если

$$p_k(i_1, \ldots, i_k) = p_{i_1} p_{i_2} \ldots p_{i_k}, \tag{3.17}$$

то соотношения (3.12), (3.14) и (3.16) выполняются, так что существует мера P на $\mathscr{F}$, удовлетворяющая условию (3.11)



0,5

Рис. 1.

и сохраняющаяся при преобразовании T. В силу эргодической теоремы асимптотическая относительная частота, с которой знак i встречается в разложении точки ω, равна p_i с точностью до множества меры 0 (в смысле меры P). Если $p_i \equiv 1/r$, то $F(x) = x$, P — мера Лебега и мы находимся в условиях примера 3.4.

Займемся теперь случаем $r = 3$, и пусть $p_0 = 1/2$, $p_1 = 0$, $p_2 = 1/2$. Тогда, согласно определению (3.13), F — функция Кантора; соответствующую меру P будем называть *мерой Кантора*. Множество точек ω, содержащих в своем разложении по основанию 3 цифры 0, 1, 2 в относительных предельных пропорциях 1/2, 0, 1/2, имеет меру Кантора, рав-

ную 1, и меру Лебега, равную 0. Таким образом, функция Кантора сингулярна[1]).

Совершенно аналогичным образом убеждаемся, что если F строится с помощью функций (3.17), то она сингулярна, если только p_i не равны тождественно $1/r$. Если ни одна из p_i не равна 0, то F — строго возрастающая функция. На рис. 1 представлен график функции F в случае $r = 2$, $p_0 = 0{,}7$, $p_1 = 0{,}3$. Если эта функция выглядит лучше функции Кантора, то это только потому, что график ее вычерчивается с точностью до толщины линии. Части графика, соответствующие отрезкам $[0, 1/2]$ и $[1/2, 1]$, идентичны с точностью до масштаба вертикальной оси, и каждая из этих частей с той же оговоркой идентична всему графику. Именно этим свойством воспроизведения[2]) должна обладать функция F, для того чтобы знаки диадического разложения $x_1(\omega)$, $x_2(\omega)$, ... образовывали процесс Бернулли с соответствующей мерой P.

Теорема существования

Специальный случай теоремы существования Колмогорова, нужный для одностороннего сдвига (пример 3.2), следует из уже рассмотренного, так как любая мера P на интервале может быть перенесена на пространство последовательностей элементов множества $\rho = \{0, 1, \ldots, r-1\}$ с помощью отображения $\omega \to (x_1(\omega), x_2(\omega), \ldots)$. Построив меру на произведении $\rho \times \rho \times \ldots$, можно, конечно, заменить ρ любым множеством из r элементов — пространство состояний совсем не обязано иметь своими элементами знаки разложения по основанию r. Нарушение предельного соотношения (3.14) соответствует в $\rho \times \rho \times \ldots$ точке с положительной вероятностью. Так как нетрудно построить точечные массы в этом пространстве, то условие (3.14) можно опустить.

Это рассуждение показывает, что если функции p_k на последовательностях длины k элементов из ρ удовлетворяют

---

[1]) Носителем вероятностной меры P является такое множество A, что $P(A) = 1$. Две вероятностные меры сингулярны одна относительно другой или взаимно сингулярны, если у них существуют непересекающиеся носители. Функция распределения сингулярна, если соответствующая мера сингулярна относительно меры Лебега.

[2]) Это напоминает мальчика, изображенного на обложке журнала рассматривающим эту обложку, и т. д. до бесконечности, только здесь он видит себя на каждой из двух смежных внутренних страниц.

4 Зак. 1491

условиям (1.3), то существует вероятностная мера P на $\rho \times \rho \times \ldots$, такая, что

$$p_k(i_1, \ldots, i_k) = P\{x_1(\omega) = i_1, \ldots, x_k(\omega) = i_k\},$$

где x_l — координатные переменные[1]). Если верно соотношение (1.4), то мера P сохраняется при одностороннем сдвиге.

Этот результат можно распространить на двусторонний сдвиг следующим образом. Пусть ψ отображает пространство $\rho \times \rho \times \ldots$ на пространство $\ldots \times \rho \times \rho \times \rho \times \ldots$ по формуле

$$\psi(\omega_1, \omega_2, \ldots) = (\ldots, \omega_4, \omega_2, \omega_1, \omega_3, \omega_5, \ldots),$$

где ω_1 стоит на нулевом месте в образе последовательности. Заданные функции p_k, удовлетворяющие условиям (1.3) и (1.4), определяют новые функции

$$p_k^+(i_1 \ldots, i_k) = \begin{cases} p_k(i_{k-1}, \ldots, i_4, i_2, i_1, i_3, \ldots, i_k) & \text{для нечетного } k, \\ p_k(i_k, \ldots, i_4, i_2, i_1, i_3, \ldots, i_{k-1}) & \text{для четного } k. \end{cases}$$

Для функций p_k^+ выполняется (1.3) (но (1.4) может не выполняться), и, следовательно, существует мера P^+ на $\rho \times \rho \times \ldots$, такая, что

$$p_k^+(i_1, \ldots, i_k) = P^+\{x_1(\omega) = i_1, \ldots, x_k(\omega) = i_k\}.$$

Для меры $P^+\psi^{-1}$ на $\ldots \times \rho \times \rho \times \rho \times$ выполняется соотношение (1.2), и она сохраняется при двустороннем сдвиге. Таким образом, мы получили доказательство того специального случая теоремы существования Колмогорова, который встретился нам в примере 1.2.

---

[1]) Действительно, если условие (3.14) не выполняется, то существуют такие последовательности $i_1, i_2, \ldots i_k$ ($i_k \neq r-1$), что

$$\lim_{l \to \infty} p_{k+l}\Big(i_1, \ldots, i_k, \underbrace{r-1, \ldots, r-1}_{l}\Big) = q_k(i_1, \ldots, i_k) > 0.$$

Заменим $p_n(i_1, \ldots, i_n)$ на

$$p_n'(i_1, \ldots, i_n) = p_n - \sum_{i_{n+1}, \ldots, i_{n+l}} q_{n+l}(i_1, \ldots, i_n, i_{n+1}, \ldots, i_{n+l}).$$

Тогда система чисел p_n' будет удовлетворять первым двум условиям (3.12) и условию (3.14). По этой системе можно построить меру на $\rho \times \rho \times \ldots$, которая, однако, не будет нормированной. Остается теперь каждой точке $\omega = (i_1, \ldots, i_k, r-1, \ldots)$, $i_k \neq r-1$, приписать меру, равную $q(i_1, \ldots, i_k)$. — *Прим. ред.*

Эргодичность и экстремальные точки

Рассмотрим вероятностную меру, сохраняемую некоторым фиксированным измеримым преобразованием T на некотором фиксированном измеримом пространстве $(\Omega, \mathcal{F})$. Назовем меру P эргодической, если относительно этой меры эргодично преобразование T. Мы покажем, что если меры P_1 и P_2 эргодические, то они либо совпадают, либо взаимно сингулярны. (Частный случай этого общего утверждения мы использовали при доказательстве сингулярности функции Кантора.) В самом деле, если P_1 и P_2 не совпадают, то $P_1(A) \neq P_2(A)$ для некоторого A из $\mathcal{F}$; если A_i — множество точек ω таких, что

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} I_A(T^k \omega) = P_i(A),$$

то A_1 и A_2 не пересекаются и $P_1(A_1) = P_2(A_2) = 1$, так что P_1 и P_2 взаимно сингулярны.

Отсюда следует, что если P — эргодическая мера, а P_1 абсолютно непрерывна относительно P (и, следовательно, тоже эргодична), то меры эти совпадают, так как они не могут быть взаимно сингулярны. Если P не эргодична, определим меру P_1 равенством $P_1(B) = P(B \mid A)$, где A — инвариантное множество и $0 < P(A) < 1$. Тогда T сохраняет P_1, а P_1 отлична от P и абсолютно непрерывна относительно нее. Таким образом, P эргодична в том и только том случае, если не существует такой меры P_1 (сохраняющейся при преобразовании T), которая отличалась бы от P и была бы относительно нее абсолютно непрерывна.

Предположим теперь, что P является взвешенной средней двух мер P_1 и P_2. Иными словами, пусть $P = \alpha_1 P_1 + \alpha_2 P_2$ (т. е. $P(B) = \alpha_1 P_1(B) + \alpha_2 P_2(B)$ для всех B из $\mathcal{F}$), где $\alpha_1 \alpha_2 > 0$, $\alpha_1 + \alpha_2 = 1$ и $P_1 \neq P_2$. Тогда P_1 отлична от P и абсолютно непрерывна относительно нее, и, значит, P не эргодична. А если так, то существует нетривиальное инвариантное множество A, так что $P(B) = P(A) P(B \mid A) + P(A^c) P(B \mid A^c)$ дает представление $P = \alpha_1 P_1 + \alpha_2 P_2$. Итак, P эргодична в том и только том случае, если она не может быть представлена в виде взвешенного среднего двух вероятностных мер, сохраняющихся преобразованием T. (Эргодическая мера P может быть, разумеется, взвешенным средним вероятностных мер, *не* сохраняющихся преобразованием T; см. пример 3.3).

Конечные меры на $\mathcal{F}$ образуют линейное пространство; вероятностные меры, сохраняющиеся при преобразовании T, образуют выпуклое множество в этом пространстве. Мы показали, что эргодические меры являются как раз экстремальными точками этого выпуклого множества.

Замечание. Функционалы от цепей Маркова дают примеры сдвигов, которые обладают свойством перемешивания, не будучи марковскими сдвигами; см., например, Розенблат [1]. Более подробные сведения, касающиеся примера 3.5, можно найти в работе Харриса [1].

4. ПРИМЕНЕНИЕ К НЕПРЕРЫВНЫМ ДРОБЯМ *

Преобразование

Любое число ω из единичного интервала можно представить в виде простой непрерывной дроби [1])

$$\omega = \frac{1\,|}{|\,a_1} + \frac{1\,|}{|\,a_2} + \ldots, \tag{4.1}$$

где элементы a_n — положительные целые числа. Это представление конечно в том и только том случае, если ω рационально.

Для изучения эргодических свойств знаков разложения числа ω по основанию r используется r-адическое преобразование $T\omega = r\omega \pmod 1$ на единичном интервале (пример 3.4) — знаки разложения $T\omega$ являются сдвинутыми на одно место знаками разложения ω. Существует преобразование, аналогичным образом сдвигающее элементы a_n, которое может быть использовано для изучения их эргодических свойств.

Пусть ω имеет представление (4.1), а другая точка единичного интервала ω' имеет представление

$$\omega' = \frac{1\,|}{|\,a_2} + \frac{1\,|}{|\,a_3} + \ldots .$$

Тогда

$$\omega = \frac{1}{a_1 + \omega'},$$

так что число $1/\omega$ имеет своей целой частью $[1/\omega]$ элемент a_1, а дробной частью $\{1/\omega\}$ — число ω'. Будем изучать преобразование, переводящее ω в $\{1/\omega\}$.

---

[1]) Об элементарных свойствах непрерывных дробей см., например, Харди и Райт [1] или начало книги Хинчина [4].

Пусть пространство Ω есть полуинтервал $[0, 1)$, а поле $\mathscr{F}$ состоит из его борелевских подмножеств. Определим преобразование T формулой

$$T\omega = \begin{cases} \{1/\omega\}, & \text{если } \omega \neq 0, \\ 0, & \text{если } \omega = 0. \end{cases}$$

Если

$$a(\omega) = \begin{cases} [1/\omega], & \text{если } \omega \neq 0, \\ \infty, & \text{если } \omega = 0, \end{cases}$$

и

$$a_n(\omega) = a\left(T^{n-1}\omega\right), \quad n = 1, 2, \ldots,$$

то $a_1(\omega)$, $a_2(\omega)$, ... являются как раз элементами непрерывной дроби, представляющей ω [1]).

Нам потребуются некоторые результаты, касающиеся непрерывных дробей. Определим целозначные функции $p_n(\omega)$, $q_n(\omega)$ рекуррентными формулами:

$$p_{-1}(\omega) = 1,\ p_0(\omega) = 0,\ p_n(\omega) = a_n(\omega)\, p_{n-1}(\omega) + p_{n-2}(\omega), \quad n \geqslant 1;$$
$$(4.2)$$
$$q_{-1}(\omega) = 0,\ q_0(\omega) = 1,\ q_n(\omega) = a_n(\omega)\, q_{n-1}(\omega) + q_{n-2}(\omega), \quad n \geqslant 1.$$

(В рациональном случае $p_n(\omega)$ и $q_n(\omega)$ определены только до тех пор, пока $a_n(\omega)$ конечны.) Стандартные рассуждения, использующие индукцию, показывают, что

$$\omega = \frac{1\ \ |}{|\, a_1(\omega)} + \ldots + \frac{1\ \ |}{|\, a_{n-1}(\omega)} + \frac{1\ \ |}{|\, a_n(\omega) + T^n\omega}, \tag{4.3}$$

$$p_{n-1}(\omega)\, q_n(\omega) - p_n(\omega)\, q_{n-1}(\omega) = (-1)^n, \ n \geqslant 0, \tag{4.4}$$

и

$$\frac{1\ \ |}{|\, a_1(\omega)} + \ldots + \frac{1\ \ |}{|\, a_{n-1}(\omega)} + \frac{1\ \ |}{|\, a_n(\omega) + t} =$$
$$= \frac{p_n(\omega) + t p_{n-1}(\omega)}{q_n(\omega) + t q_{n-1}(\omega)}, \ n \geqslant 1, \ 0 \leqslant t \leqslant 1. \tag{4.5}$$

Положив в (4.5) $t = 0$, получаем формулу для n-го приближения:

$$\frac{1\ \ |}{|\, a_1(\omega)} + \ldots + \frac{1\ \ |}{|\, a_n(\omega)} = \frac{p_n(\omega)}{q_n(\omega)}.$$

Соотношения (4.3) и (4.5) дают

$$\omega = \frac{p_n(\omega) + (T^n\omega)\, p_{n-1}(\omega)}{q_n(\omega) + (T^n\omega)\, q_{n-1}(\omega)},$$

---

[1]) $T^n\omega = 0$ для некоторого n в том и только том случае, если ω рационально. Условимся считать элементы $a_n(\omega)$ бесконечными при $n \geqslant k$, если наше представление — конечная k-членная непрерывная дробь.

и, используя (4.4), получаем

$$\left|\omega - \frac{p_n(\omega)}{q_n(\omega)}\right| = \frac{1}{q_n(\omega)\left((T^n\omega)^{-1} q_n(\omega) + q_{n-1}(\omega)\right)}.$$

Так как

$$a_{n+1}(\omega) \leqslant (T^n\omega)^{-1} \leqslant a_{n+1}(\omega) + 1,$$

то

$$\frac{1}{q_n(\omega)\left(q_n(\omega) + q_{n+1}(\omega)\right)} \leqslant \left|\omega - \frac{p_n(\omega)}{q_n(\omega)}\right| \leqslant \frac{1}{q_n(\omega)\, q_{n+1}(\omega)} \quad (n \geqslant 1). \tag{4.6}$$

Наконец, нам потребуется неравенство

$$\left|\ln \frac{\omega}{p_n(\omega)/q_n(\omega)}\right| \leqslant \frac{1}{2^{n-2}}, \quad n \geqslant 1. \tag{4.7}$$

Справедливость его при $n = 1$ может быть установлена непосредственной проверкой. Так как по соображениям индукции

$$p_n(\omega) \geqslant 2^{(n-2)/2}, \quad q_n(\omega) \geqslant 2^{(n-1)/2}, \quad n \geqslant 2, \tag{4.8}$$

то

$$\left|\frac{\omega}{p_n(\omega)/q_n(\omega)} - 1\right| \leqslant \frac{1}{2^{n-1}}, \quad n \geqslant 2.$$

Для $n \geqslant 2$ неравенство (4.7) следует из последнего неравенства.

Пусть a_k, $k = 1, \ldots, n$, — положительные целые числа [1]). Обозначим символом $\Delta_{a_1, \ldots, a_n}$ множество точек ω таких, что $a_1(\omega) = a_1, \ldots, a_n(\omega) = a_n$. Множество $\Delta a_1 \ldots a_n$, которое мы будем называть фундаментальным интервалом ранга n, играет здесь ту же роль, что r-адический интервал при изучении разложений по основанию r. Множество $\Delta_{a_1 \ldots a_n}$ является образом полуинтервала $[0, 1)$, полученным с помощью функции $\psi_{a_1 \ldots a_n}$, определенной равенством

$$\psi_{a_1 \ldots a_n}(t) = \frac{1|}{|a_1} + \ldots + \frac{1|}{|a_{n-1}} + \frac{1|}{|a_n + t}, \quad 0 \leqslant t < 1.$$

Из самого вида функции ясно, что $\psi_{a_1 \ldots a_n}$ убывает при нечетном n и возрастает при четном n. В силу (4.5) имеем

$$\psi_{a_1 \ldots a_n}(t) = \frac{p_n + t p_{n-1}}{q_n + t q_{n-1}}, \tag{4.9}$$

---

[1]) Не путайте с функциями $a_k(\omega)$; чтобы избежать недоразумений, мы не будем опускать аргумент ω при записи этих функций (а также функций $p_k(\omega)$ и $q_k(\omega)$).

где p_n и q_n определяются через элементы a_k рекуррентными соотношениями, подобными (4.2). Поэтому

$$\Delta_{a_1 \dots a_n} = \begin{cases} \left[\dfrac{p_n}{q_n}, \dfrac{p_n + p_{n-1}}{q_n + q_{n-1}}\right), & \text{если } n \text{ четно,} \\ \left[\dfrac{p_n + p_{n-1}}{q_n + q_{n-1}}, \dfrac{p_n}{q_n}\right), & \text{если } n \text{ нечетно.} \end{cases}$$

Из соотношения (4.4) следует

$$\lambda\left(\Delta_{a_1 \dots a_n}\right) = \frac{1}{q_n (q_n + q_{n-1})}, \tag{4.10}$$

где λ означает меру Лебега. Таким образом, фундаментальные интервалы ранга n образуют разбиение пространства Ω на интервалы длины, не большей 2^{-n+1}. В частности, класс всех фундаментальных интервалов порождает σ-поле $\mathscr{F}$ борелевских множеств.

Мера Гаусса

Легко видеть, что преобразование T не сохраняет меру Лебега λ. Однако на $\mathscr{F}$ существует мера, которая, как мы увидим, сохраняется при преобразовании T, а именно мера Гаусса

$$P(A) = \frac{1}{\ln 2} \int\limits_A \frac{dx}{1+x}, \quad A \in \mathscr{F}. \tag{4.11}$$

Так как P и λ абсолютно непрерывны относительно друг друга, то соответствующие им множества меры 0 совпадают. Таким образом, если последовательность элементов $a_1(\omega)$, $a_2(\omega)$, ... обладает некоторым свойством почти всюду относительно P, то она обладает этим свойством и почти всюду относительно λ.

Для того чтобы доказать, что T сохраняет P, достаточно показать, что оно сохраняет меры отрезков $[0, \alpha]$. Так как

$$T^{-1}[0, \alpha] = \bigcup_{k=1}^{\infty} \left[\frac{1}{k+\alpha}, \frac{1}{k}\right],$$

то нужно проверить только выполнение равенства

$$\int\limits_0^{\alpha} \frac{dx}{1+x} = \sum_{k=1}^{\infty} \int\limits_{1/(k+\alpha)}^{1/k} \frac{dx}{1+x}.$$

Равенство справедливо, поскольку k-й член в правой его части равен

$$\ln\left(1+\frac{1}{k}\right)-\ln\left(1+\frac{1}{k+\alpha}\right)=$$
$$=\ln\left(1+\frac{\alpha}{k}\right)-\ln\left(1+\frac{\alpha}{k+1}\right)=\int\limits_{\alpha/(k+1)}^{\alpha/k}\frac{dx}{x+1}.$$

Докажем теперь, что T эргодично относительно P. Фиксируем $a_1 \ldots a_n$ и будем обозначать $\omega_{a_1 \ldots a_n}$ через ψ и $\Delta_{a_1 \ldots a_n}$ через Δ_n. Длина интервала Δ_n равна $\pm(\psi(1)-\psi(0))$, и если $0 \leqslant x < y \leqslant 1$, то длина интервала

$$\{\omega : x \leqslant T^n\omega < y\} \cap \Delta_n$$

равна $\pm(\psi(y)-\psi(x))$ со знаком $+$ или $-$ в зависимости от того, четно или нечетно n. Поэтому

$$\lambda\left(T^{-n}[x,\ y) \mid \Delta_n\right)=\frac{\psi(y)-\psi(x)}{\psi(1)-\psi(0)}.$$

В силу (4.9) и (4.4) имеем

$$\lambda\left(T^{-n}[x,\ y) \mid \Delta_n\right)=(y-x)\frac{q_n(q_n+q_{n-1})}{(q_n+xq_{n-1})(q_n+yq_{n-1})}. \tag{4.12}$$

Так как второй множитель в правой части равенства лежит между 1/2 и 2, то

$$\frac{1}{2}\lambda(A) \leqslant \lambda\left(T^{-n}A \mid \Delta_n\right) \leqslant 2\lambda(A), \tag{4.13}$$

где $A=[x,\ y)$. Тогда (4.13) справедливо и в случае, если A — объединение непересекающихся интервалов, следовательно, справедливо для любого A из $\mathscr{F}$.

Так как плотность меры Гаусса, определенной выражением (4.11), всегда находится между $1/2\ln 2$ и $1/\ln 2$, то

$$\frac{\lambda(M)}{2\ln 2} \leqslant P(M) \leqslant \frac{\lambda(M)}{\ln 2},\quad M \in \mathscr{F}. \tag{4.14}$$

Из (4.13) следует, что

$$\frac{1}{C}P(A) \leqslant P\left(T^{-n}A \mid \Delta_n\right) \leqslant CP(A) \tag{4.15}$$

для всех A из $\mathscr{F}$, где C — абсолютная константа ($C=4/\ln 2$).

Предположим, что A — инвариантное множество. Тогда $C^{-1}P(A) \leqslant P(A \mid \Delta_n)$ и, следовательно, если $P(A)>0$, то $C^{-1}P(\Delta_n) \leqslant P(\Delta_n \mid A)$. Поэтому неравенство

$$\frac{1}{C}P(E) \leqslant P(E \mid A) \tag{4.16}$$

выполняется для конечных объединений E непересекающихся фундаментальных интервалов. Поскольку эти множества образуют поле, порождающее $\mathcal{F}$, то (4.16) верно для любого E из $\mathcal{F}$. Если положить $E = A^c$, то получаем, что $P(A)$ должна равняться 1. Поэтому T эргодично относительно P.

Из эргодической теоремы следует, что если f — интегрируемая функция на единичном интервале, то

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k\omega) = \frac{1}{\ln 2} \int_0^1 \frac{f(x)}{1+x}\, dx \text{ п. в.} \tag{4.17}$$

Здесь не имеет значения, к какой мере, P или λ, относятся слова „интегрируема" и „п. в.".

Пусть f — характеристическая функция множества $\{\omega : a_1(\omega) = k\}$. Тогда асимптотическая относительная частота, с которой число k встречается среди элементов $a_1(\omega), a_2(\omega), \ldots$, почти всюду равна

$$\frac{1}{\ln 2} \int_{1/(k+1)}^{1/k} \frac{dx}{1+x} = \frac{1}{\ln 2} \ln \frac{(k+1)^2}{k(k+2)}.$$

В частности, последовательность элементов $a_1(\omega), a_2(\omega), \ldots$ почти всюду неограниченна.

Положив $f(\omega) = \ln a_1(\omega)$, видим, что

$$\lim_{n\to\infty} \sqrt[n]{a_1(\omega) \ldots a_n(\omega)} = \prod_{k=1}^{\infty} \left(1 + \frac{1}{k^2+2k}\right)^{\ln k/\ln 2} \text{ п. в.}$$

Если $f(\omega) = a_1(\omega)$, то интеграл от f расходится к $+\infty$. Мы приходим с помощью формальных операций к предельному соотношению

$$\lim_{n\to\infty} \frac{a_1(\omega) + \ldots + a_n(\omega)}{n} = \infty \text{ п. в.}$$

С помощью усечения нетрудно доказать это строго.

В применениях к диофантовым приближениям для нас важнее величина $q_n(\omega)$, чем $a_n(\omega)$. Мы докажем, что

$$\lim_{n\to\infty} \frac{1}{n} \ln q_n(\omega) = \frac{\pi^2}{12 \ln 2} \text{ п. в.} \tag{4.18}$$

Покажем сначала, что

$$\frac{1}{q_n(\omega)} = \prod_{k=1}^{n} \left(\frac{1}{|a_k(\omega)} + \ldots + \frac{1}{|a_n(\omega)}\right). \tag{4.19}$$

Из рекуррентных соотношений (4.2) с помощью индукции получаем $p_{j+1}(\omega)=q_j(T\omega)$, и поэтому

$$\frac{1}{q_n(\omega)}=\prod_{k=1}^{n}\frac{p_{n+1-k}(T^{k-1}\omega)}{q_{n+1-k}(T^{k-1}\omega)}, \tag{4.20}$$

так как числитель k-го множителя в правой части сокращается со знаменателем $(k+1)$-го множителя. Но (4.20) есть как раз (4.19).

В силу (4.7) имеем

$$\left|\ln T^{k-1}\omega-\ln\left(\frac{1}{|a_k(\omega)|}+\ldots+\frac{1}{|a_n(\omega)|}\right)\right|\leqslant\frac{1}{2^{n-k-1}},\quad 1\leqslant k\leqslant n,$$

и, следовательно, в силу (4.19)

$$\ln\frac{1}{q_n(\omega)}=\sum_{k=1}^{n}\ln T^{k-1}\omega+\sum_{k=1}^{n}\frac{\theta}{2^{n-k-1}},\quad |\theta|\leqslant 1.$$

Поэтому

$$\frac{1}{n}\ln q_n(\omega)=-\frac{1}{n}\sum_{k=0}^{n-1}\ln T^k\omega+4\frac{\theta}{n},\quad |\theta|\leqslant 1. \tag{4.21}$$

В силу эргодической теоремы

$$\lim_{n\to\infty}\left\{-\frac{1}{n}\sum_{k=0}^{n-1}\ln T^k\omega\right\}=-\frac{1}{\ln 2}\int_0^1\frac{1}{1+x}\ln x\,dx\quad \text{п. в.}$$

Интегрируя по частям, получаем

$$\frac{1}{\ln 2}\int_0^1\ln(1+x)\frac{dx}{x}=\frac{1}{\ln 2}\sum_{k=0}^{\infty}(-1)^k\int_0^1\frac{x^k}{k+1}\,dx=$$

$$=\frac{1}{\ln 2}\sum_{k=0}^{\infty}\frac{(-1)^k}{(k+1)^2}=\frac{\pi^2}{12\ln 2}.$$

Таким образом, (4.18) следует из (4.21).

Соотношение (4.18) имеет несколько простых, но интересных следствий. Например, используя (4.6), получаем

$$\lim_{n\to\infty}\frac{1}{n}\ln\left|\omega-\frac{p_n(\omega)}{q_n(\omega)}\right|=-\frac{\pi^2}{6\ln 2}\quad \text{п. в.}$$

Таким образом, расхождение между ω и его n-м приближением почти всюду порядка $e^{-n\pi^2/6\ln 2}$. Далее, если $\Delta_n(\omega)$ —

фундаментальный интервал ранга n, содержащий ω, то в силу (4.10)

$$\lim_{n\to\infty} \frac{1}{n} \ln \lambda(\Delta_n(\omega)) = -\frac{\pi^2}{6\ln 2} \text{ п. в.}$$

Наконец, в силу (4.17) имеем

$$\lim_{n\to\infty} \frac{1}{n} \ln P(\Delta_n(\omega)) = -\frac{\pi^2}{6\ln 2} \text{ п. в.} \tag{4.22}$$

Применение к диофантовым приближениям

Пусть $\{\alpha_n\}$ — последовательность положительных чисел и E_n означает событие $\{a_n(\omega) > \alpha_n\}$. Так как величина $P(E_n) = P\{a_n(\omega) > \alpha_n\}$ порядка $1/\alpha_n$, то из леммы Бореля — Кантелли следует, что если $\sum 1/\alpha_n$ сходится, то число наступлений события $a_n(\omega) > \alpha_n$ конечно, за исключением, быть может, множества меры 0 (относительно P или λ).

Предположим, что $\sum 1/\alpha_n$ расходится. В силу (4.15)

$$P(E_{n+1} | \Delta_n) \geqslant \frac{1}{C'(\alpha_{n+1}+1)}$$

для любого фундаментального интервала ранга n, где C' — абсолютная константа. Поэтому

$$P\left(E^c_m \cap \ldots \cap E^c_{m+k}\right) \leqslant \prod_{i=0}^{k}\left(1 - \frac{1}{C'(\alpha_{m+1}+1)}\right).$$

Если $\sum 1/\alpha_n$ расходится, то произведение стремится к 0 при $k\to\infty$; следовательно,

$$P\left(E^c_m \cap E^c_{m+1} \cap \ldots\right) = 0.$$

Так как это верно для каждого m, то число наступлений события $a_n(\omega) > \alpha_n$ бесконечно, за исключением, быть может, множества меры 0. Мы можем теперь сформулировать следующий результат.

Теорема 4.1. *Число наступлений события $a_n(\omega) > \alpha_n$ бесконечно с вероятностью* 0 *или* 1 *в зависимости от того, сходится или расходится* $\sum 1/\alpha_n$.

Этот результат и соотношение (4.18) приводят к следующей фундаментальной теореме метрической теории аппроксимации:

Теорема 4.2. *Пусть $f(q)$ — положительная функция натурального аргумента q.* (а) *Если $qf(q)$ — невозрастающая*

функция и $\sum_q f(q) = \infty$, *то для почти всех* ω *неравенство*

$$\left|\omega - \frac{p}{q}\right| < \frac{f(q)}{q} \tag{4.23}$$

имеет бесконечно много решений в целых p *и* q. (b) *Если* $\sum_q f(q) < \infty$, *то для почти всех* ω *неравенство* (4.23) *имеет не более чем конечное число решений.*

Доказательство. (a) Выберем и зафиксируем натуральное N (скажем, $N = 4$), такое, что $\ln N > \pi^2/12 \ln 2$. Из (4.18) следует, что неравенство

$$\frac{1}{n} \ln q_n(\omega) < \ln N \tag{4.24}$$

выполняется с точностью до множества меры 0 для всех, кроме конечного числа, значений n. Если $\varphi(n) = N^n f(N^n)$, то, поскольку $qf(q)$ не возрастает, имеем

$$\sum_{q=N^n}^{N^{n+1}-1} f(q) \leqslant 12 \ln N \cdot \varphi(n),$$

так что $\sum_n \varphi(n)$ расходится, если расходится $\sum_q f(q)$. Из теоремы 4.1 вытекает, что неравенство

$$a_{n+1}(\omega) > \frac{1}{\varphi(n)} \tag{4.25}$$

с точностью до множества меры 0 справедливо для бесконечно многих n.

Если выполнено неравенство (4.25), то в силу (4.6) и (4.2)

$$\left|\omega - \frac{p_n(\omega)}{q_n(\omega)}\right| < \frac{1}{q_n(\omega)\, q_{n+1}(\omega)} \leqslant \frac{1}{a_{n+1}(\omega)\, q_n(\omega)^2} \leqslant \frac{\varphi(n)}{q_n(\omega)^2}.$$

Но если (4.24) также выполнено, так что $q_n(\omega) < N^n$, то в силу того, что $qf(q)$ не возрастает, имеем

$$\varphi(n) = N^n f(N^n) \leqslant q_n(\omega) f(q_n(\omega)),$$

так что

$$\left|\omega - \frac{p_n(\omega)}{q_n(\omega)}\right| < \frac{f(q_n(\omega))}{q_n(\omega)}.$$

Так как неравенства (4.24) и (4.25) с точностью до множества точек ω меры 0 выполняется одновременно для бесконечного множества значений n, то часть (a) теоремы 4.2 доказана.

Часть (b) вытекает из простых свойств меры Лебега. Если H_q — множество точек ω, для которых неравенство (4.23) выполняется при каком-нибудь натуральном p, то H_q является объединением интервалов длины $2f(q)/q$ с центрами в точках вида p/q. Так как таких точек на единичном интервале только q, то

$$\lambda(H_q) \leqslant 2f(q).$$

Таким образом, часть (b) следует из леммы Бореля — Кантелли.

Перемешивание и проблема Гаусса

Усиление нашего доказательства эргодичности преобразования T дает нам новые сведения.

Пусть $\mathscr{G}_n$ есть σ-поле, порожденное множествами вида $\{\omega: a_k(\omega) = a\}$ при $k \geqslant n$, и пусть $\mathscr{G}_\infty = \bigcap_{n=1}^{\infty} \mathscr{G}_n$. Множество из σ-поля $\mathscr{G}_\infty$, которое мы назовем *хвостовым* [1]) σ-полем, зависит только от „бесконечно далекого будущего". Покажем сначала, что любое инвариантное множество „почти" лежит в этом хвостовом σ-поле, т. е. если A инвариантно, то $P(A+B) = 0$ для некоторого $B \in \mathscr{G}_\infty$. Так как $A \in \mathscr{F}$, то $T^{-n}A \in \mathscr{G}_n$ даже для неинвариантного A. Но если A строго инвариантно, то $A = T^{-n}A \in \mathscr{G}_n$ для всех n и, следовательно, $A \in \mathscr{G}_\infty$. Если же A просто инвариантно, то $P(A+B) = 0$ для некоторого строго инвариантного множества B.

Так как в $\mathscr{G}_\infty$ существуют множества, не являющиеся строго инвариантными [2]), то, возможно, $\mathscr{G}_\infty$ содержит множество A, для которого $0 < P(A) < 1$, даже если T эргодично. (Если бы T не было эргодичным, то $\mathscr{G}_\infty$ содержало бы такое множество A, так как инвариантные множества „почти" лежат в $\mathscr{G}_\infty$.) Если $P(A) > 0$, то для любого n множество A имеет вид $A = T^{-n}B$, где $B \in \mathscr{F}$. Но в силу (4.15)

$$\frac{1}{C} P(A) = \frac{1}{C} P\left(T^{-n}B\right) = \frac{1}{C} P(B) \leqslant P\left(T^{-n}B \mid \Delta_n\right) = P(A \mid \Delta_n).$$

для любого фундаментального интервала Δ_n ранга n. Так как $P(A) > 0$, то $C^{-1}P(\Delta_n) \leqslant P(\Delta_n \mid A)$. Как и раньше [см. (4.16)], отсюда следует, что $P(A) = 1$.

---

[1]) В отечественной литературе принят термин „остаточное σ-поле". — *Прим. ред.*

[2]) Например, множество точек ω, для которых $a_n(\omega) = 1$ при бесконечном множестве четных значений n.

Таким образом, хвостовое σ-поле $\mathscr{G}_\infty$ тривиально в том смысле, что оно содержит только множества меры 0 или 1, — условие не менее сильное, чем эргодичность. В действительности, как мы увидим позже (в конце § 11), из этого условия можно сделать вывод о перемешивающем свойстве преобразования T.

В письме к Лапласу[1]) Гаусс утверждал, что

$$\lim_{n\to\infty} \lambda\{\omega:\ T^n\omega < x\} = \frac{\ln(1+x)}{\ln 2} = P\,[0,\ x) \tag{4.26}$$

для каждого x из единичного интервала, и интересовался оценкой ошибки при использовании n-го приближения. Запишем (4.26) в виде

$$\lim_{n\to\infty} \int\limits_{T^{-n}A} (\ln 2)(1+\omega)\,P\,(d\omega) = P\,(A), \tag{4.27}$$

где $A = [0,\ x)$. Из того, что преобразование T является перемешивающим относительно меры P, вытекает, что

$$\lim_{n\to\infty} \int\limits_{T^{-n}A} f(\omega)\,P\,(d\omega) = P\,(A) \int\limits_0^1 f(\omega)\,P\,(d\omega)$$

для любой характеристической функции f и любого борелевского множества A. Так как функция $(\ln 2)(1+\omega)$ равномерно аппроксимируема ступенчатыми функциями, то имеет место (4.27) (для любого борелевского множества A). Таким образом, утверждение Гаусса вытекает из того, что преобразование T является перемешивающим.

Другие методы показывают, что перемешивание равномерно экспоненциально, т. е.

$$P\left(T^{-(n+k)}A \mid \Delta_k\right) = P\,(A)\left(1+\theta\rho^n\right), \tag{4.28}$$

где $|\theta| \leqslant K$, K и ρ — положительные константы ($\rho < 1$), не зависящие от A, n, k и Δ_k. Отсюда следует, что сходимость в (4.26) равномерна и экспоненциальна.

З а м е ч а н и е. Многие из приведенных результатов, включая теорему 4.2, доказаны Хинчиным [1,2]. Его доказательства осложнены тем, что он не использует эргодической теоремы. Дёблин [1, стр. 336], по-видимому, первым доказал (4.17) во всей общности, применяя эргодическую теорему.

---

[1]) Это письмо цитируется Успенским [1]. Не ясно, каким доказательством этого утверждения располагал Гаусс.

Другое доказательство эргодичности преобразования T см. в работе Рыль-Нарджевского [1].

Кузьмин [1] первым доказал (4.26), он получил оценку ошибки аппроксимации порядка $\rho^{\sqrt{n}}$. Леви [1, 2] улучшил ее до ρ^n, доказав (4.28). Дёблину [1] принадлежит другое доказательство соотношения (4.28) и еще много вероятностных результатов относительно непрерывных дробей.

Существует класс теоретико-числовых преобразований единичного интервала, содержащий в виде специальных случаев преобразование непрерывных дробей и преобразование $\omega \to r\omega \pmod 1$; см. Реньи [1] и Рохлин [3].

Кац [1] исследовал различные связи между теорией вероятностей и другими областями математики.

ГЛАВА 2

Энтропия

5. ПРОБЛЕМА ИЗОМОРФИЗМА

Существуют такие пары сохраняющих меру преобразований, которые, будучи формально различными, по существу совпадают. Сдвиг Бернулли ни в каком существенном отношении не изменится, если элементы пространства состояний ρ обозначить по-новому. Очевидно также, что вращение единичной окружности $T\omega = c\omega$ (пример 1.5) не отличается от преобразования $T\omega = \omega + (\arg c)/2\pi \pmod 1$ единичного интервала с мерой Лебега, а диадическое преобразование $T\omega = 2\omega \pmod 1$ единичного интервала не отличается от преобразования $T\omega = \omega^2$ единичной окружности с мерой Лебега.

Несколько более глубокий пример дает сравнение преобразования $T\omega = 2\omega \pmod 1$ на единичном интервале и одностороннего сдвига Бернулли с пространством состояний $\rho = \{0, 1\}$, $p_0 = p_1 = 1/2$. Если мы сопоставим точки единичного интервала и элементы прямого произведения $\rho \times \rho \times \ldots$ с помощью соответствия

$$0, \omega_1\omega_2 \ldots \longleftrightarrow (\omega_1, \omega_2, \ldots)$$

(мы на время игнорируем неоднозначность диадических разложений), то множества соответствующих точек (например, левый полуинтервал и цилиндр $\{\omega: x_1(\omega) = 0\}$) имеют одинаковую меру и оба преобразования действуют по существу одинаковым образом, ибо одно переводит точку $0, \omega_1\omega_2 \ldots$ в $0, \omega_2\omega_3 \ldots$, а другое — точку $(\omega_1, \omega_2, \ldots)$ в $(\omega_2, \omega_3, \ldots)$.

Изоморфизм

Необходимо ввести понятие изоморфизма для сохраняющих меру преобразований, подобное понятию изоморфизма, скажем, для групп. Пусть T и $\tilde{T}$ — сохраняющие меру преобразования, определенные на вероятностных пространствах $(\Omega, \mathscr{F}, P)$ и $(\tilde{\Omega}, \tilde{\mathscr{F}}, \tilde{P})$ соответственно[1]). В качестве предва-

[1]) Далее всюду, где специально не оговорено противное, предполагается, что все преобразования T, $\tilde{T}$ и т. д. сохраняют меру.

рительного определения возьмем следующее: преобразования T и $\tilde{T}$ *изоморфны*, если существует отображение φ пространства Ω *на* $\tilde{\Omega}$ такое, что 1) φ взаимно однозначно; 2) если $\tilde{A}=\varphi A$, то $A \in \mathcal{F}$ в том и только том случае, когда $\tilde{A} \in \tilde{\mathcal{F}}$, при этом $P(A)=\tilde{P}(\tilde{A})$; 3) равенство $\varphi T\omega = \tilde{T}\varphi\omega$ выполняется для всех ω. Условия 1 и 2 требуют, чтобы отображение φ сохраняло структуру измеримых пространств $(\Omega, \mathcal{F}, P)$ и $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$. Условие 3 требует, чтобы φ переводило T в $\tilde{T}$: один и тот же результат получается независимо от того, каким из двух путей совершается переход от верхнего левого Ω к нижнему правому $\tilde{\Omega}$ в диаграмме

$$\begin{tikzcd} \Omega \arrow[r, "T"] \arrow[d, "\varphi"'] & \Omega \arrow[d, "\varphi"] \\ \tilde{\Omega} \arrow[r, "\tilde{T}"] & \tilde{\Omega} \end{tikzcd}$$

Пары преобразований, рассмотренные в начале этого параграфа, изоморфны в указанном смысле, но все же данное определение неудовлетворительно. Предположим, что T — тождественное преобразование пространства Ω, состоящего из одной точки, а $\tilde{T}$ — тождественное преобразование пространства $\tilde{\Omega}$, состоящего из двух точек с массами 0 и 1 ($\tilde{\mathcal{F}}$ состоит из всех четырех подмножеств). Хотя отображения φ, подобного описанному выше, не существует уже потому, что пространства Ω и $\tilde{\Omega}$ имеют разные мощности, преобразования T и $\tilde{T}$ по существу одинаковы — точка пространства $\tilde{\Omega}$ с массой 0 не должна идти в счет. Вследствие неоднозначности диадических разложений для приведенного выше соответствия

$$0, \omega_1\omega_2 \ldots \longleftrightarrow (\omega_1, \omega_2, \ldots)$$

возникает аналогичная трудность.

Нам нужно определение, которое было бы нечувствительно к множествам меры 0, т. е. такое определение, которое признавало бы преобразования T и $\tilde{T}$ изоморфными, если они становятся изоморфными после выбрасывания множества меры 0 из одного или из обоих пространств Ω и $\tilde{\Omega}$. Пусть теперь Ω_0 есть то, что остается после удаления из Ω некоторого множества меры 0, т. е. Ω_0 — множество из $\mathcal{F}$ меры 1. Тогда T может рассматриваться как преобразование, определенное лишь на Ω_0, в том и только том

случае, если оно переводит Ω_0 в себя: $T\omega$ принадлежит Ω_0 всякий раз, как ω принадлежит Ω_0, или, что то же самое, $\Omega_0 \supset T\Omega_0$ (или, что то же, $\Omega_0 \subset T^{-1}\Omega_0$). Будем теперь считать T и $\tilde{T}$ изоморфными, если, будучи суженными до преобразований, действующих на подмножествах Ω_0 и $\tilde{\Omega}_0$, удовлетворяющих соотношениям $\Omega_0 \subset T^{-1}\Omega_0$ и $\tilde{\Omega}_0 \subset \tilde{T}^{-1}\tilde{\Omega}_0$, они становятся изоморфными в смысле предварительного определения. Это приводит к следующему определению.

Пусть существуют множества Ω_0 *из* $\mathscr{F}$ *и* $\tilde{\Omega}_0$ *из* $\tilde{\mathscr{F}}$, *имеющие меру* 1, *и отображение* φ *множества* Ω_0 *на* $\tilde{\Omega}_0$, *обладающие следующими свойствами.*

(I_1) *Отображение* φ *взаимно однозначно.*

(I_2) *Если* $A \subset \Omega_0$ *и* $\tilde{A} = \varphi A$, *то* $A \in \mathscr{F}$ *в том и только том случае, когда* $\tilde{A} \in \tilde{\mathscr{F}}$, *при этом* $P(A) = \tilde{P}(\tilde{A})$.

(I_3) *Имеют место соотношения*

$$\Omega_0 \subset T^{-1}\Omega_0 \ (\text{т. е. } \Omega_0 \supset T\Omega_0), \tag{5.1}$$

$$\tilde{\Omega}_0 \subset \tilde{T}^{-1}\tilde{\Omega}_0 \ (\text{т. е. } \tilde{\Omega}_0 \supset \tilde{T}\tilde{\Omega}_0) \tag{5.2}$$

и, наконец,

$$\varphi T\omega = \tilde{T}\varphi\omega \tag{5.3}$$

выполняется для любого ω *из* Ω_0.

В этом случае мы говорим, что преобразования T *и* $\tilde{T}$ *изоморфны* (*точнее, изоморфны* $(\Omega, \mathscr{F}, P, T)$ *и* $(\tilde{\Omega}, \tilde{\mathscr{F}}, \tilde{P}, \tilde{T})$)[1]. Для того чтобы подчеркнуть роль множеств Ω_0, $\tilde{\Omega}_0$ и отображения φ, можно говорить, что преобразования T и $\tilde{T}$ изоморфны *относительно тройки* $(\Omega_0, \tilde{\Omega}_0, \varphi)$.

(Условия (I_1) и (I_2) представляют собой определение одинаковости самих пространств с мерой $(\Omega, \mathscr{F}, P)$ и $(\tilde{\Omega}, \tilde{\mathscr{F}}, \tilde{P})$. Условие ($I_3$) — только в нем фигурируют преобразования T и $\tilde{T}$ — утверждает, что эти преобразования абстрактно тождественны.)

Следующие замечания содержат утверждения теории меры относительно нульмерных множеств, нужные для точной трактовки понятия изоморфизма.

---

[1]) Это понятие иногда называют изоморфизмом по модулю 0, или изоморфизмом по модулю множеств меры 0, или почти изоморфизмом. Отказавшись от предварительного определения, можно опустить подобные уточнения.

Замечание 1. Данное выше предварительное определение соответствует частному случаю, когда можно выбрать тройку $(\Omega_0, \tilde{\Omega}_0, \varphi)$ таким образом, что $\Omega_0 = \Omega$ и $\tilde{\Omega}_0 = \tilde{\Omega}$. Тогда равенство

$$T^{-n}\varphi^{-1}\tilde{A} = \varphi^{-1}\tilde{T}^{-n}\tilde{A} \tag{5.4}$$

верно для любого подмножества $\tilde{A}$ множества $\tilde{\Omega}$ и любого $n \geqslant 0$.

Замечание 2. В связи с соотношением (5.3) заметим, что если $\omega \in \Omega_0$, то $\varphi\omega \in \tilde{\Omega}_0$, так что в силу (5.1) и (5.2) $T\omega$ и $\tilde{T}\varphi\omega$ принадлежат множествам Ω_0 и $\tilde{\Omega}_0$ соответственно. В этом случае условие (5.3) требует соответствия точек $T\omega$ и $\tilde{T}\varphi\omega$ относительно отображения φ. Ослабим условие (I_3), опуская требования (5.1) и (5.2) и предполагая только справедливость (5.3) всякий раз, как элементы ω и $T\omega$ оба принадлежат множеству Ω_0. Если тройка $(\Omega_0, \tilde{\Omega}_0, \varphi)$ удовлетворяет условиям определения, где (I_3) ослаблено указанным образом, то преобразования T и $\tilde{T}$ не обязательно изоморфны относительно тройки $(\Omega_0, \tilde{\Omega}_0, \varphi)$, но легко видеть, что они изоморфны относительно тройки $(\Omega_1, \tilde{\Omega}_1, \varphi_1)$, где $\Omega_1 = \bigcap_{n=0}^{\infty} T^{-n}\Omega_0$, $\tilde{\Omega}_1 = \varphi\Omega_1$ и φ_1 — сужение φ на Ω_1. (На самом деле преобразования T и $\tilde{T}$ изоморфны, даже если вместо условия (I_3) мы потребуем, чтобы (5.3) выполнялось всегда, когда имеют место соотношения $\omega \in \Omega_0$, $T\omega \in \Omega_0$ и $\tilde{T}\varphi\omega \in \tilde{\Omega}_0$.)

Замечание 3. Мы должны показать, что изоморфизм является отношением эквивалентности. Очевидно, что преобразование T изоморфно самому себе; следовательно, изоморфизм рефлексивен. Если T изоморфно $\tilde{T}$ относительно тройки $(\Omega_0, \tilde{\Omega}_0, \varphi)$, то $\tilde{T}$ изоморфно T относительно тройки $(\tilde{\Omega}_0, \Omega_0, \varphi^{-1})$; следовательно, изоморфизм симметричен. Для доказательства транзитивности изоморфизма нужно показать, что если T изоморфно $\tilde{T}$ и $\tilde{T}$ изоморфно третьему преобразованию T' (определенному на пространстве $(\Omega', \mathcal{F}', P')$), то T изоморфно T'. Допустим, что T изоморфно $\tilde{T}$ относительно тройки $(\Omega_0, \tilde{\Omega}_0, \varphi)$ и $\tilde{T}$ изоморфно T' относительно тройки $(\tilde{\Omega}_1, \Omega'_1, \psi)$. Пусть $\tilde{\Omega}_2 = \tilde{\Omega}_0 \cap \tilde{\Omega}_1$, $\Omega_2 = \varphi^{-1}\tilde{\Omega}_2$, $\Omega'_2 = \psi\tilde{\Omega}_2$, и пусть φ_2 и ψ_2 являются сужениями отображений φ и ψ на Ω_2 и $\tilde{\Omega}_2$ соответственно. Нетрудно показать, что преобразование T

изоморфно $\tilde{T}$ относительно тройки $(\Omega_2, \tilde{\Omega}_2, \varphi_2)$ и $\tilde{T}$ изоморфно T' относительно тройки $(\tilde{\Omega}_2, \Omega_2', \psi_2)$. Так как образ множества Ω_2 при отображении φ_2 совпадает с множеством $\tilde{\Omega}_2$, на котором определено отображение ψ_2, то с помощью композиции φ_2 и ψ_2 можно получить взаимно однозначное отображение ξ множества Ω_2 на Ω_2': $\xi(\omega) = \psi_2(\varphi_2(\omega))$. Теперь нетрудно убедиться, что T и T' изоморфны относительно тройки $(\Omega_2, \Omega_2', \xi)$.

Замечание 4. Если преобразование T изоморфно $\tilde{T}$ относительно тройки $(\Omega_0, \tilde{\Omega}_0, \varphi)$, то в силу соотношений (5.1) и (5.2)

$$\Omega_0 \subset T^{-1}\Omega_0 \subset T^{-2}\Omega_0 \subset \ldots \left(\text{т. е. } \Omega_0 \supset T\Omega_0 \supset T^2\Omega_0 \supset \ldots\right) \quad (5.5)$$

и

$$\tilde{\Omega}_0 \subset \tilde{T}^{-1}\tilde{\Omega}_0 \subset \tilde{T}^{-2}\tilde{\Omega}_0 \subset \ldots \ (\text{т. е. } \tilde{\Omega}_0 \supset \tilde{T}\tilde{\Omega}_0 \supset \tilde{T}^2\tilde{\Omega}_0 \supset \ldots). \quad (5.6)$$

Далее, из (5.3) следует по индукции, что если $\omega \in \Omega_0$, то

$$\varphi T^n \omega = \tilde{T}^n \varphi\omega \quad (5.7)$$

выполняется при всех $n \geqslant 0$. (В частности, T^n и $\tilde{T}^n$ изоморфны при всех n.) Наконец, из соотношений (5.5), (5.6) и (5.7) следует, что если $\tilde{A}$ и $\tilde{B}$ — множества из поля $\tilde{\mathscr{F}}$ и если

$$A = \varphi^{-1}(\tilde{\Omega}_0 \cap \tilde{A}), \quad B = \varphi^{-1}(\tilde{\Omega}_0 \cap \tilde{B}), \quad (5.8)$$

то

$$A \cap T^{-n}B = \varphi^{-1}\left(\tilde{\Omega}_0 \cap \tilde{A} \cap \tilde{T}^{-n}\tilde{B}\right), \quad n \geqslant 0, \quad (5.9)$$

откуда

$$P\left(A \cap T^{-n}B\right) = \tilde{P}\left(\tilde{A} \cap \tilde{T}^{-n}\tilde{B}\right), \quad (5.10)$$

ибо множества Ω_0 и $\tilde{\Omega}_0$ имеют меру 1.

Замечание 5. Обратимые изоморфные преобразования T и $\tilde{T}$ фактически изоморфны относительно тройки $(\Omega_0, \tilde{\Omega}_0, \varphi)$, обладающей тем специальным свойством, что множества Ω_0 и $\tilde{\Omega}_0$ строго инвариантны: $\Omega_0 = T^{-1}\Omega_0$ и $\tilde{\Omega}_0 = \tilde{T}^{-1}\tilde{\Omega}_0$ (если это необходимо, заменим Ω_0 на $\bigcap\limits_{n=-\infty}^{\infty} T^n\Omega_0$ и $\tilde{\Omega}_0$ на $\bigcap\limits_{n=-\infty}^{\infty} \tilde{T}^n\tilde{\Omega}_0$). Однако, вообще говоря, это не справедливо: возьмем, например, в качестве Ω пространство, состоящее из одной точки, а в качестве преобразования $\tilde{T}$

односторонний сдвиг, при котором вся масса сосредоточена в некоторой точке вида $(i, i, i, \ldots)$.

Различные пары преобразований, которые мы обсуждали в начале этого параграфа, изоморфны в смысле нашего определения. Рассмотрим еще несколько таких пар.

Пример 5.1. Преобразуем единичный квадрат Ω, удваивая каждую координату по оси x и вдвое уменьшая каждую координату по оси y. Преобразованный квадрат состоит из прямоугольников B и C (рис. 2). Переведем

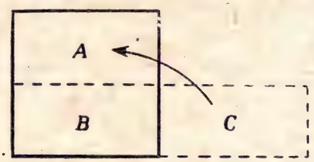


Рис. 2.

теперь прямоугольник C в прямоугольник A с помощью сдвига. Мы определили таким образом некоторое преобразование T, переводящее пространство Ω в себя и сохраняющее меру Лебега. Это преобразование называют преобразованием пекаря. Между преобразованием T и сдвигом Бернулли (1/2, 1/2)[1]) существует изоморфизм, осуществляемый посредством соответствия

$$(x, y) = (0, x_1x_2, \ldots, 0, y_1y_2, \ldots) \longleftrightarrow (\ldots, y_2, y_1, x_1, x_2, \ldots),$$

где $0, x_1x_2 \ldots$ и $0, y_1y_2 \ldots$ — двоичное разложение чисел x и y.

Пример 5.2. Пусть преобразование T — сдвиг Маркова с пространством состояний ρ, стационарными вероятностями p_i и вероятностями перехода p_{ij}. Тогда преобразование T^2 изоморфно сдвигу Маркова $\tilde{T}$ с пространством состояний $\tilde{\rho} = \rho^2$, стационарными вероятностями $p_{(ij)} = p_ip_{ij}$ и вероятностями перехода $p_{(i,\,j)\,(k,\,l)} = p_{jk}p_{kl}$. Если $\Omega\,[\tilde{\Omega}]$ — пространство бесконечных в обе стороны последовательностей элементов пространства $\rho\,[\tilde{\rho}]$, а $x_n\,[\tilde{x}_n]$ — координатные переменные, то нужно только положить $\Omega_0 = \Omega$, $\tilde{\Omega}_0 = \tilde{\Omega}$ и

[1]) Под сдвигом Бернулли $(p_1, \ldots, p_r)$ понимают сдвиг с вероятностями p_i на пространстве состояний ρ. Природа элементов пространства ρ совершенно безразлична; вектор вероятностей записывают символом $(p_1, \ldots, p_r)$, хотя пространство ρ может и не состоять из первых r целых чисел. Везде далее, если не оговорено противное, все сдвиги будем считать двусторонними.

определить отображение φ при помощи равенства $\tilde{x}_n(\varphi\omega) = (x_{2n}(\omega),\ x_{2n+1}(\omega))$.

Пример 5.3. Если в качестве стационарных вероятностей и вероятностей перехода, соответствующих преобразованию $\tilde{T}$ из предыдущего примера, взять соответственно $p_{(i,\,j)} = p_i p_{ij}$ и $p_{(i,\,j)\,(k,\,l)} = \delta_{jk} p_{kl}$, то преобразование $\tilde{T}$ будет изоморфно самому преобразованию T. Мы снова полагаем $\Omega_0 = \Omega$, но $\tilde{\Omega}_0$ определяем на этот раз как множество точек $\tilde{\omega}$, для которых вторая компонента координаты $\tilde{x}_n(\tilde{\omega})$ совпадает с первой компонентой координаты $\tilde{x}_{n+1}(\tilde{\omega})$ при всех n, а отображение φ определяем соотношением $\tilde{x}_n(\varphi\omega) = (x_{n-1}(\omega),\ x_n(\omega))$.

Пример 5.4. Пусть преобразование T — двусторонний сдвиг с пространством состояний $\rho = \{0, 1\}$; предположим, что наша мера P такова, что не существует точечных масс. Пусть $\tilde{\Omega}$ пространство бесконечных в обе стороны последовательностей действительных чисел, как в примере 1.7. Пусть образ $\varphi\omega$ элемента ω пространства Ω является точкой пространства $\tilde{\Omega}$, n-я координата которой есть

$$\sum_{k=0}^{\infty} x_{n-k}(\omega)/2^{k+1} = 0,\ x_n(\omega)\,x_{n-1}(\omega)\ \ldots \quad \text{(двоичное разложение).}$$

Тогда сдвиг T изоморфен сдвигу $\tilde{T}$, определенному на пространстве $\tilde{\Omega}$ с мерой $\tilde{P} = P\varphi^{-1}$. Можно показать, что в пространстве $\tilde{\Omega}$ координатные переменные $\{\tilde{x}_n\}$ образуют относительно меры $\tilde{P}$ марковский процесс с единичным интервалом в качестве пространства состояний.

Следующий пример показывает, что наше определение изоморфизма все еще не вполне удовлетворительно.

Пример 5.5. Сравним тождественное преобразование T, определенное на пространстве Ω, состоящем из одной точки, с тождественным преобразованием $\tilde{T}$ на пространстве $\tilde{\Omega}$, состоящем из двух точек, причем поле $\tilde{\mathcal{F}}$ содержит только пустое множество и само пространство $\tilde{\Omega}$. Хотя преобразования T и $\tilde{T}$ неизоморфны, они имеют по существу одинаковую структуру, ибо поле $\tilde{\mathcal{F}}$ таково, что у нас нет возможности различать две точки пространства $\tilde{\Omega}$.

Этот пример характеризует некоторую трудность, которую можно обойти, заменяя понятие изоморфизма понятием сопряженности. Грубо говоря, на этом пути игнорируют в зна-

чительной степени точки и точечные преобразования и имеют дело с множествами и преобразованиями множеств, для которых формулируют соответствующие понятия одинаковости. Два преобразования, совпадающие в смысле этой теории, называют сопряженными.

Вопрос состоит в том, какое понятие одинаковости принять. Преобразования T и $\tilde{T}$ примера 5.5 удовлетворяют определению сопряженности. Тот факт, что эти преобразования неизоморфны, можно приписать дефекту вероятностного пространства $(\tilde{\Omega}, \tilde{\mathscr{F}}, \tilde{P})$, на котором определено преобразование $\tilde{T}$. Понятия изоморфизма и сопряженности совпадают для большинства естественных пространств; так как эти понятия совпадают и для всех специфических пространств, встречающихся в этой книге в связи с примерами (за исключением только что рассмотренного), то мы выберем изоморфизм в качестве нашего понятия одинаковости сохраняющих меру преобразований. Однако дальше в этом параграфе мы все же определим понятие сопряженности и докажем, что оно действительно совпадает с понятием изоморфизма для преобразований, определенных на некотором классе пространств, достаточно широком, чтобы включать в себя все те конкретные пространства, к которым мы применяем наши общие результаты.

Инварианты

Так же как можно задаться вопросом, являются ли знакопеременная группа из пяти букв и группа симметрий икосаэдра различными конкретными представлениями одной и той же алгебраической структуры — т. е. являются ли они изоморфными в смысле теории групп (а это так и есть), — можно задать вопрос, являются ли два конкретных сохраняющих меру преобразования абстрактно одинаковыми в смысле принятого определения изоморфизма. Например, изоморфен ли сдвиг Бернулли (1/2, 1/2) сдвигу Бернулли (1/3, 1/3, 1/3)?

Для доказательства изоморфности двух преобразований нужно построить соответствующую тройку $(\Omega_0, \tilde{\Omega}_0, \varphi)$. Для доказательства неизоморфности двух преобразований нужно отвергнуть — предпочтительно одновременно — все возможные варианты выбора тройки $(\Omega_0, \tilde{\Omega}_0, \varphi)$. Следующее рассуждение показывает, что так же как две группы не могут быть изоморфными, если одна из них коммутативна, а другая нет, так и два сохраняющих меру преобразования не могут

быть изоморфными, если одно из них обладает свойством перемешивания, а другое этим свойством не обладает.

Предположим, что T и $\tilde{T}$ изоморфны относительно тройки $(\Omega_0, \tilde{\Omega}_0, \varphi)$ и T обладает свойством перемешивания; мы должны доказать, что тогда $\tilde{T}$ также необходимо обладает этим свойством. Для того чтобы уловить основной момент рассуждения, предположим сначала, что $\Omega_0 = \Omega$ и $\tilde{\Omega}_0 = \tilde{\Omega}$ (так что преобразования T и $\tilde{T}$ изоморфны в смысле предварительного суженного определения). Мы имеем тогда для любых множеств $\tilde{A}$ и $\tilde{B}$ поля $\tilde{\mathscr{F}}$ (см. (5.4))

$$\tilde{P}(\tilde{A} \cap \tilde{T}^{-n}\tilde{B}) = P(\varphi^{-1}\tilde{A} \cap \varphi^{-1}\tilde{T}^{-n}\tilde{B}) =$$
$$= P(\varphi^{-1}\tilde{A} \cap T^{-n}\varphi^{-1}\tilde{B}) \to P(\varphi^{-1}\tilde{A})\, P(\varphi^{-1}\tilde{B}) = \tilde{P}(\tilde{A})\, \tilde{P}(\tilde{B}),$$

где использовано свойство перемешивания преобразования T. Таким образом, преобразование $\tilde{T}$ перемешивающее.

Некоторое изменение этого доказательства покрывает общий случай, когда не предполагается, что $\Omega_0 = \Omega$ и $\tilde{\Omega}_0 = \tilde{\Omega}$. Для множеств $\tilde{A}$ и $\tilde{B}$ поля $\tilde{\mathscr{F}}$ положим $A = \varphi^{-1}(\tilde{\Omega}_0 \cap \tilde{A})$ и $B = \varphi^{-1}(\tilde{\Omega}_0 \cap \tilde{B})$; в силу (5.10) имеем

$$\tilde{P}(\tilde{A} \cap \tilde{T}^{-n}\tilde{B}) = P(A \cap T^{-n}B) \to P(A)\, P(B) = \tilde{P}(\tilde{A})\, \tilde{P}(\tilde{B}).$$

Итак, свойство перемешивания является *инвариантом*: если одно из пары сохраняющих меру преобразований обладает этим свойством, то и другое им обладает. Используя этот инвариант, убеждаемся, что сдвиг Бернулли не может быть изоморфен вращению окружности или сдвигу Маркова с периодической матрицей вероятностей перехода.

Другим инвариантом является эргодичность (нужно заменить в приведенном доказательстве обычные пределы пределами по Чезаро и применить теорему 1.4). Поэтому, например, сдвиг Маркова с приводимой матрицей вероятностей перехода не может быть изоморфен никакому сдвигу Маркова с неприводимой матрицей[1]).

Обратимость не является вполне инвариантным свойством: изменим обратимое преобразование, отображая все точки некоторого множества меры 0 в одну и ту же точку. Если, однако, преобразование T изоморфно некоторому обра-

[1]) Инвариантом более мощным, чем перемешивание и эргодичность, является спектральная структура изометрического оператора, порожденного преобразованием T. Этот инвариант будет рассмотрен в конце настоящего параграфа.

тимому преобразованию $\tilde{T}$, то T взаимно однозначно на некотором подмножестве Ω_0 меры 1. Отсюда следует, что двусторонний сдвиг не может быть изоморфен одностороннему, если только вся мера, отвечающая последнему, не сосредоточена в точках вида $(i, i, i, \ldots)$. Вообще говоря, можно сформулировать инвариантное определение обратимости, но мы не станем этого делать.

Чем больше структур различает инвариант, тем он полезнее; лучше всего, если он полный. Размерность является полным инвариантом в векторных пространствах (скажем, вещественных) в том смысле, что два векторных пространства одинаковой размерности необходимо изоморфны. Эргодичность не является полным инвариантом для сохраняющих меру преобразований: очевидно, что существуют неизоморфные пары преобразований, каждое из которых эргодично (или оба не эргодичны). Свойство перемешивания также не является полным инвариантом.

Вернемся к вопросу об изоморфизме сдвигов Бернулли $(1/2, 1/2)$ и $(1/3, 1/3, 1/3)$. Эргодичность и перемешивание не являются, разумеется, инвариантами достаточно сильными, чтобы различать эти сдвиги, ибо оба сдвига перемешивающие и, следовательно, эргодические. Вопрос о том, изоморфны ли эти два сдвига, много лет не поддавался решению. Наконец Колмогоров разрешил эту проблему в отрицательном смысле, введя новый численный инвариант — *энтропию* сохраняющего меру преобразования. Оказалось, что сдвиги Бернулли $(1/2, 1/2)$ и $(1/3, 1/3, 1/3)$ имеют различные энтропии и потому неизоморфны. Инвариант Колмогорова — это существенным образом измененное понятие энтропии, введенное ранее Шенноном в теории информации. Здесь мы нарушим историческую последовательность. В этой главе мы изучим детально инвариант Колмогорова, а позднее применим полученные результаты к теории информации.

Энтропия

Нам потребуется несколько определений, которые мы дадим сначала формально, отложив мотивировку на дальнейшее. Везде далее мы будем обозначать буквами $\mathscr{A}$, $\mathscr{B}$ и $\mathscr{C}$ конечные подполя поля $\mathscr{F}$. (Конечное подполе автоматически является σ-полем.) Если $\{A_1, \ldots, A_n\}$ есть $\mathscr{F}$-разбиение пространства Ω, т. е. конечная совокупность непересекающихся непустых элементов поля $\mathscr{F}$, объединение которых совпадает со всем пространством Ω, то класс всех конечных объединений элементов этого разбиения является

конечным подполем поля $\mathscr{F}$. Легко видеть, что и обратно, любое конечное подполе получается таким образом из некоторого $\mathscr{F}$-разбиения. Элементы такого разбиения будут называться *атомами* соответствующего конечного поля. Между $\mathscr{F}$-разбиениями и конечными подполями поля $\mathscr{F}$ существует полная двойственность. Например, $\mathscr{A} \subset \mathscr{B}$ в том и только том случае, если разбиение, соответствующее конечному подполю $\mathscr{B}$, мельче, чем разбиение, соответствующее $\mathscr{A}$, в том смысле, что каждый атом подполя $\mathscr{A}$ является объединением атомов подполя $\mathscr{B}$.

Для любой совокупности множеств $\mathscr{E}$ можно записать $T^{-n}\mathscr{E} = \{T^{-n}E : E \in \mathscr{E}\}$; если $\mathscr{E}$ — поле (или σ-поле, или конечное поле, или $\mathscr{F}$-разбиение), то то же самое справедливо и по отношению к $T^{-n}\mathscr{E}$. Если преобразование T обратимо, это же верно и для $T^{n}\mathscr{E} = \{T^{n}E : E \in \mathscr{E}\}$. Если $\mathscr{E}_\alpha$, $\alpha \in A$, — произвольные совокупности множеств, то символом $\bigvee\limits_{\alpha \in A} \mathscr{E}_\alpha$ будем обозначать σ-поле, порожденное объединением $\bigcup\limits_{\alpha \in A} \mathscr{E}_\alpha$ $\left(\text{в конечном случае будем писать } \bigvee\limits_{i=1}^{n} \mathscr{E}_i = \mathscr{E}_1 \vee \ldots \vee \mathscr{E}_n\right)$. Если $\mathscr{A}$ и $\mathscr{B}$ — конечные поля, то $\mathscr{A} \vee \mathscr{B}$ — также конечное поле. В самом деле, атомы поля $\mathscr{A} \vee \mathscr{B}$ являются по предыдущему (непустыми) пересечениями $A \cap B$ атомов A поля $\mathscr{A}$ и атомов B поля $\mathscr{B}$.

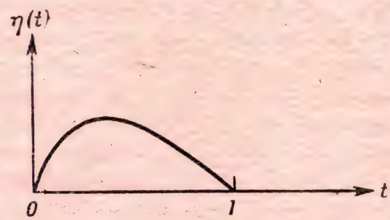


Рис. 3.

Далее во всей книге будем обозначать символом $\eta(t)$ функцию, определенную на единичном интервале формулой

$$\eta(t) = \begin{cases} -t \ln t, & \text{если } 0 < t \leqslant 1, \\ 0 = 0 \ln 0, & \text{если } t = 0. \end{cases} \tag{5.11}$$

Мы постоянно будем использовать основные свойства функции η: она неотрицательна, непрерывна, строго выпукла и $\eta(0) = \eta(1) = 0$ (рис. 3).

Энтропия преобразования T определяется в три шага. Энтропия конечного поля $\mathcal{A}$ определяется формулой

$$H(\mathcal{A}) = \sum_A \eta(P(A)) = -\sum_A P(A)\ln P(A), \tag{5.12}$$

где суммирование производится по атомам A поля $\mathcal{A}$. Энтропия конечного поля $\mathcal{A}$ относительно преобразования T есть

$$h(\mathcal{A}, T) = \varlimsup_{n\to\infty} \frac{1}{n} H\left(\bigvee_{k=0}^{n-1} T^{-k}\mathcal{A}\right). \tag{5.13}$$

(Оказывается, что верхний предел в этом выражении совпадает с обычным пределом.) Наконец, энтропия преобразования T задается выражением

$$h(T) = \sup h(\mathcal{A}, T), \tag{5.14}$$

где верхняя грань берется по всем конечным подполям $\mathcal{A}$ поля $\mathcal{F}$.

Для понимания интуитивных идей, лежащих в основе этих определений, рассмотрим кость с r гранями. В качестве меры количества *случайности* при однократном бросании этой кости возьмем неотрицательную величину

$$\sum_{i=1}^{r} \eta(p_i) = -\sum_{i=1}^{r} p_i \ln p_i, \tag{5.15}$$

где $p_1, \ldots, p_r$ — вероятности, соответствующие различным граням кости. Здесь мы не будем углубляться в подробности, но, вообще говоря, можно вывести (5.15) из некоторой системы аксиом, которым должна была бы удовлетворять мера случайности. Заметим, что выражение (5.15) достигает максимума (равного $\ln r$) в том и только том случае, если [1]) каждая вероятность p_i равна $1/r$; подобная кость интуитивно представляется „наиболее случайной“. Другого экстремума — нуля — выражение (5.15) достигает в том и только том случае, когда одна из вероятностей p_i равна 1, а другие 0; такая кость — „наименее случайная“.

В любом случае мы рассматриваем выражение (5.15) как меру количества случайности в эксперименте, состоящем в однократном бросании нашей кости, и называем его *энтропией*

---

[1]) Функция $\sum_{i=1}^{r} \eta(p_i + t\varepsilon_i)$ имеет отрицательную вторую производную по действительной переменной t; следовательно, (5.15) есть строго выпуклая функция вектора вероятностей $(p_1, \ldots, p_r)$. Максимум ее можно найти дифференцированием.

этого эксперимента. Эта величина измеряет также количество *неопределенности*, содержащейся в этом эксперименте, т. е. количество неопределенности до бросания кости относительно того, каков будет его результат. Наконец, эта величина измеряет *информацию*, содержащуюся в этом эксперименте, или количество информации, получаемой в результате бросания. Тот факт, что случайность и неопределенность имеют естественную общую меру, неудивителен. Вследствие „формулы“

прирост информации = устраненная неопределенность

представляется разумным, что неопределенность и информация должны измеряться с помощью одной и той же функции. (Предполагается, что экспериментатор знает вероятности p_i, но не знает заранее, какая грань выпадает при каждом отдельном бросании.)

Конечное поле $\mathscr{A}$ играет роль эксперимента с конечным числом исходов. Богиня Тихе выбирает точку ω из пространства Ω в соответствии с вероятностной мерой P, но открывает экспериментатору только тот атом поля $\mathscr{A}$, который содержит точку ω. Атомы поля $\mathscr{A}$ выступают в качестве исходов этого эксперимента, а выражение (5.12) измеряет информацию (а также неопределенность), в нем содержащуюся.

Например, для того чтобы выбрать некоторую точку ω из пространства двенадцати точек

$$\Omega\colon\ \begin{array}{cccccc} (\text{Г},\ 1) & (\text{Г},\ 2) & (\text{Г},\ 3) & (\text{Г},\ 4) & (\text{Г},\ 5) & (\text{Г},\ 6) \\ (\text{Р},\ 1) & (\text{Р},\ 2) & (\text{Р},\ 3) & (\text{Р},\ 4) & (\text{Р},\ 5) & (\text{Р},\ 6), \end{array}$$

Тихе бросает монету со сторонами Г и Р и кость с гранями 1, 2, 3, 4, 5, 6. Если нам известен лишь результат бросания монеты, то мы знаем только, из какой строки выбрана точка ω; эти две строки являются атомами конечного поля, формально представляющего эксперимент, состоящий в бросании монеты. Энтропия этого конечного поля не зависит от индивидуальных вероятностей данных двенадцати точек; она зависит только от вероятностей двух указанных атомов.

Так как атомы поля $\mathscr{A} \vee \mathscr{B}$ представляют собой пересечения атомов поля $\mathscr{A}$ и поля $\mathscr{B}$, то, если нам известно, какой атом поля $\mathscr{A} \vee \mathscr{B}$ содержит точку ω, то тем самым мы знаем, какой атом поля $\mathscr{A}$ и какой атом поля $\mathscr{B}$ содержат ω. Таким образом, поле $\mathscr{A} \vee \mathscr{B}$, рассматриваемое как эксперимент, является соединением экспериментов, соответствующих полям $\mathscr{A}$ и $\mathscr{B}$.

Если T — сохраняющее меру преобразование, то $T^{-1}\mathcal{A}$ — конечное поле с тем же числом атомов, что и поле $\mathcal{A}$, и соответствующие атомы имеют равные вероятности. Так как поля $\mathcal{A}$ и $T^{-1}\mathcal{A}$ имеют одинаковую вероятностную структуру, то их можно считать реализациями одного и того же эксперимента. Отметим, что эти реализации не обязательно независимы; знание того, что точка ω лежит в некотором определенном атоме поля $\mathcal{A}$, может помочь экспериментатору определить, какой атом поля $T^{-1}\mathcal{A}$ ее содержит (т. е. какой атом поля $\mathcal{A}$ содержит точку $T\omega$). Будем рассматривать поля $\mathcal{A}$ и $T^{-1}\mathcal{A}$ как реализации эксперимента, совершающиеся в последовательные моменты времени (скажем, в следующие друг за другом дни), и такое представление мы сохраним даже и в случае, когда преобразование T не является сдвигом.

В этой интерпретации конечное поле $\bigvee\limits_{k=0}^{n-1} T^{-k}\mathcal{A}$ соответствует сложному эксперименту, состоящему из n реализаций $\mathcal{A}, T^{-1}\mathcal{A}, \ldots, T^{-(n-1)}\mathcal{A}$ эксперимента, соответствующего полю $\mathcal{A}$. Для того чтобы получить информацию, приходящуюся на одну реализацию эксперимента, разделим на n количество информации $H\left(\bigvee\limits_{k=0}^{n-1} T^{-k}\mathcal{A}\right)$ сложного эксперимента. При $n \to \infty$ информация $n^{-1}H\left(\bigvee\limits_{k=0}^{n-1} T^{-k}\mathcal{A}\right)$, приходящаяся на одну реализацию, стремится (как мы увидим) к пределу $h(\mathcal{A}, T)$, который мы рассматриваем как среднюю скорость создания информации при большом числе реализаций эксперимента $\mathcal{A}$. Функция $h(T)$ есть верхняя грань этих скоростей по всем экспериментам $\mathcal{A}$.

Рассмотрим в качестве специального случая сдвиг. Пусть пространство состояний ρ есть некоторый алфавит, а координатные переменные x_n — буквы этого алфавита, последовательно доставляемые некоторым *источником информации*. В английском языке буква E встречается значительно чаще, чем Q, как это известно всем, включая изобретателя азбуки Морзе, который обозначил букву E символом $\cdot$, а Q символом — — $\cdot$ —. Если развивать далее эту идею, то она приведет к плодотворной точке зрения, согласно которой английский текст можно воспроизвести с помощью вероятностного механизма. Сдвиг дает подходящую математическую модель, в которой мера P описывает структуру языка.

Пусть теперь $\mathcal{A}$ — конечное поле с атомами $\{\omega : x_0(\omega) = i\}$, где i пробегает алфавит ρ. Это поле, которое мы будем называть полем событий, наблюдавшихся в момент времени 0, ибо оно определяет букву, полученную в момент времени 0, имеет энтропию $\sum_{i\in\rho}\eta(P\{x_0 = i\})$, измеряющую количество информации, доставляемой источником при создании им одной отдельной буквы. Так как $T^{-k}\{x_0 = i\} = \{x_k = i\}$, то атомами поля $\bigvee_{k=0}^{n-1} T^{-k}\mathcal{A}$ являются r^n множеств $\{x_0 = i_0, \ldots$ $\ldots, x_{n-1} = i_{n-1}\}$. Таким образом,

$$\frac{1}{n} H\left(\bigvee_{k=0}^{n-1} T^{-k}\mathcal{A}\right) = \frac{1}{n} \sum_{i_0 \ldots i_{n-1}} \eta(P\{x_0 = i_0, \ldots, x_{n-1} = i_{n-1}\}) \tag{5.16}$$

есть информация, приходящаяся на букву в сообщении длины n. Предел $h(\mathcal{A}, T)$ есть информационная скорость источника [1]).

Если T — сдвиг Бернулли $(p_1, \ldots, p_r)$, то простые вычисления преобразуют (5.16) в

$$\frac{1}{n} \sum_{i_0 \ldots i_{n-1}} \eta\left(p_{i_0} p_{i_1} \ldots p_{i_{n-1}}\right) = -\sum_i p_i \ln p_i,$$

так что источник, последовательно доставляющий буквы независимым образом, имеет информационную скорость

$$h(\mathcal{A}, T) = -\sum_i p_i \ln p_i.$$

Заметим, что эта скорость наибольшая, когда все вероятности p_i, соответствующие различным буквам, равны $1/r$, т. е. когда мы находимся в „чисто случайной“ ситуации, что кажется сначала парадоксальным. Но чем ближе источник к чисто случайному, тем менее он стереотипен и, следовательно, более информативен. Несущественно, что именно создает источник информации, важно только, насколько предсказуемо то, что он создает.

Возвращаясь к математике, покажем, что энтропия инвариантна по отношению к изоморфизму. Пусть T изо-

---

[1]) Это определение Шеннона. Идея Колмогорова состояла в изучении функции $h(T)$, которая определена для любого преобразования T (а не только для сдвигов) и используется в эргодической теории.

морфно $\tilde{T}$ относительно тройки $(\Omega_0, \tilde{\Omega}_0, \varphi)$, где для простоты мы сначала предположим, что $\Omega_0 = \Omega$ и $\tilde{\Omega}_0 = \tilde{\Omega}$. Тогда каждому конечному подполю $\mathcal{A}$ поля $\mathcal{F}$ соответствует некоторое конечное подполе $\tilde{\mathcal{A}} = \varphi\mathcal{A} = \{\varphi A : A \in \mathcal{A}\}$ поля $\tilde{\mathcal{F}}$, и обратно. Так как отображение φ переводит меру P в $\tilde{P}$, то $H(\mathcal{A}) = H(\tilde{\mathcal{A}})$. Если конечные подполя $\mathcal{A}$ и $\tilde{\mathcal{A}}$ поставлены друг другу в соответствие, то в силу (5.4) друг другу соответствуют $\bigvee_{k=0}^{n-1} T^{-k}\mathcal{A}$ и $\bigvee_{k=0}^{n-1} T^{-k}\tilde{\mathcal{A}}$, и, следовательно, в силу (5.13) имеем $h(\mathcal{A}, T) = h(\tilde{\mathcal{A}}, \tilde{T})$. Таким образом, для каждого подполя $\mathcal{A}$ существует такое подполе $\tilde{\mathcal{A}}$, что $h(\mathcal{A}, T) = h(\tilde{\mathcal{A}}, \tilde{T})$, и обратно. Взяв верхнюю грань, видим, что $h(T) = h(\tilde{T})$.

Если не делать специального предположения, что $\Omega_0 = \Omega$ и $\tilde{\Omega}_0 = \tilde{\Omega}$, то потребуется несколько более сложное рассуждение. Так как изоморфизм — отношение симметричное, то достаточно доказать, что $h(\tilde{T}) \leqslant h(T)$. Пусть $\tilde{\mathcal{A}}$ — некоторое конечное подполе поля $\tilde{\mathcal{F}}$ с атомами $\tilde{A}_1, \ldots, \tilde{A}_r$. Положим $A_i = \varphi^{-1}(\tilde{\Omega}_0 \cap \tilde{A}_i)$, $i = 1, \ldots, r$, и пусть $\mathcal{A}$ — такое конечное подполе поля $\mathcal{F}$, которое имеет своими атомами $r+1$ множество $A_1, \ldots, A_r, \Omega_0^c$. Используя обобщение формулы (5.10), получаем

$$H\left(\bigvee_{k=0}^{n-1} \tilde{T}^{-k}\tilde{\mathcal{A}}\right) = \sum_{i_0 \ldots i_{n-1}} \eta\left(\tilde{P}\left(\bigcap_{k=0}^{n-1} \tilde{T}^{-k}\tilde{A}_{i_k}\right)\right) = \sum_{i_0 \ldots i_{n-1}} \eta\left(P\left(\bigcap_{k=0}^{n-1} T^{-k}A_{i_k}\right)\right).$$

Эта последняя сумма распространяется на все атомы $\bigvee_{k=0}^{n-1} T^{-k}\mathcal{A}$, за исключением тех, которые содержат множитель $T^{-k}\Omega_0^c$, имеющий меру 0. Поэтому $H\left(\bigvee_{k=0}^{n-1} \tilde{T}^{-k}\tilde{\mathcal{A}}\right) = H\left(\bigvee_{k=0}^{n-1} T^{-k}A\right)$, так что $h(\tilde{\mathcal{A}}, \tilde{T}) = h(\mathcal{A}, T)$. Неравенство $h(\tilde{\mathcal{A}}, \tilde{T}) \leqslant h(T)$ выполняется для любого конечного поля $\tilde{\mathcal{A}}$, следовательно $h(\tilde{T}) \leqslant h(T)$.

Итак, изоморфизм сохраняет энтропию. Отметим здесь, что энтропия не является полным инвариантом. Если T_k — циклическая перестановка k точек равной массы (σ-поле состоит из всех подмножеств этого пространства), то $h(T_k) = 0$

для всех k, хотя очевидно, что перестановки T_k для различных k неизоморфны.

Очень важно понимать, в чем состоит различие между функциями $h(\mathcal{A}, T)$ и $h(T)$ и зачем последняя вводится. Если в качестве энтропии преобразования T взята функция $h(\mathcal{A}, T)$ для некоторого „естественно" выбранного подполя $\mathcal{A}$, например в случае сдвига для поля событий, наблюдавшихся в момент времени 0, то вследствие неинвариантного определения она может оказаться бесполезной для проблемы изоморфизма. Мы определяем $h(T)$ как верхнюю грань для $h(\mathcal{A}, T)$ именно для того, чтобы сделать энтропию инвариантной. Но тогда возникает вопрос, как ее вычислять. Мы вычислили, например, $h(\mathcal{A}, T)$ для сдвига Бернулли и для поля событий, наблюдавшихся в момент времени 0, но возможно, что некоторая $h(\mathcal{B}, T)$ превосходит $h(\mathcal{A}, T)$. Если мы пожелаем вычислить энтропию самого сдвига Бернулли, потребуется нечто иное.

Колмогоров получил важный результат, который мы сейчас используем: *если T обратимо и $\mathcal{A}$ — такое конечное поле, что*[1] $\bigvee_{n=-\infty}^{\infty} T^n\mathcal{A} = \mathcal{F}$, *то* $h(T) = h(\mathcal{A}, T)$. Предполагая справедливость этой теоремы, которую вместе с ее вариантами мы докажем в § 7, можно вычислить энтропию сдвига Бернулли. В самом деле, если $\mathcal{A}$ — поле событий, наблюдавшихся в момент времени 0, то поле $\bigvee_{n=-\infty}^{\infty} T^n\mathcal{A}$ содержит все множества вида $T^{-n}\{x_0 = i\} = \{x_n = i\}$, следовательно, все цилиндры, и должно, таким образом, совпадать с σ-полем, порожденным этими цилиндрами. Следовательно, $h(T) = h(\mathcal{A}, T) = -\sum_i p_i \ln p_i$, что является не определением, а следствием нашей теории. В частности, сдвиги Бернулли $(1/2, 1/2)$ и $(1/3, 1/3, 1/3)$ имеют энтропии $\ln 2$ и $\ln 3$ соответственно и, следовательно, неизоморфны.

Энтропия является, таким образом, инвариантом достаточно сильным, чтобы решить поставленную раньше задачу. Теперь стала ясна и наша программа. Мы должны вывести свойства функций $H(\mathcal{A})$, $h(\mathcal{A}, T)$ и $h(T)$, достаточные для

---

[1] Напомним, что $\bigvee_{n=-\infty}^{\infty} T^n\mathcal{A}$ есть σ-поле, порожденное объединением $\bigcup_{n=-\infty}^{\infty} T^n\mathcal{A}$.

того, чтобы доказать результат Колмогорова, который занимает центральное место во всей этой теории.

С проблемой изоморфизма тесно связана теория кодирования, изложенная в гл. 5. Многие результаты настоящей главы находят там дальнейшее применение.

Изоморфизм и сопряженность *

Предположим, что преобразование T изоморфно $\tilde{T}$ относительно тройки $(\Omega_0, \tilde{\Omega}_0, \varphi)$. Можно использовать этот изоморфизм для того, чтобы сопоставить множества поля $\mathcal{F}$ и множества поля $\tilde{\mathcal{F}}$ следующим образом.

Будем называть множества A и B поля $\mathcal{F}$ эквивалентными и писать $A \sim B$ всякий раз, когда $P(A+B)=0$; то же относится и к множествам поля $\tilde{\mathcal{F}}$. Это понятие эквивалентности рефлексивно, симметрично и транзитивно. Пусть A — множество поля $\mathcal{F}$ и $\tilde{A}$ — множество поля $\tilde{\mathcal{F}}$; будем сопоставлять их друг другу всякий раз, когда существуют такие множества A_0 и $\tilde{A}_0$, принадлежащие полям $\mathcal{F}$ и $\tilde{\mathcal{F}}$ соответственно, что $A_0 \subset \Omega_0$, $\tilde{A}_0 \subset \tilde{\Omega}_0$, $A \sim A_0$, $\tilde{A} \sim \tilde{A}_0$ и $\tilde{A}_0 = \varphi A_0$. Такого рода сопоставление множеств A и $\tilde{A}$ (т. е. существование указанных множеств A_0 и $\tilde{A}_0$) мы будем обозначать $A \leftrightarrow \tilde{A}$.

Это сопоставление является многозначным соответствием между полями $\mathcal{F}$ и $\tilde{\mathcal{F}}$. Легко убедиться, что это соответствие обладает следующими свойствами. (Далее все множества предполагаются принадлежащими полю $\mathcal{F}$ или полю $\tilde{\mathcal{F}}$.)

(C_1) *Для любого множества A существует по крайней мере одно множество $\tilde{A}$ такое, что $A \leftrightarrow \tilde{A}$; для любого множества $\tilde{A}$ существует по крайней мере одно множество A такое, что $A \leftrightarrow \tilde{A}$.*

(C_2) *Если $A \leftrightarrow \tilde{A}$, то $A \leftrightarrow \tilde{B}$ в том и только том случае, когда $\tilde{A} \sim \tilde{B}$; если $A \leftrightarrow \tilde{A}$, то $B \leftrightarrow \tilde{A}$ в том и только том случае, когда $A \sim B$.*

(C_3) *Если $A \leftrightarrow \tilde{A}$, то $\Omega - A \leftrightarrow \tilde{\Omega} - \tilde{A}$; если $A_n \leftrightarrow \tilde{A}_n$, где множество индексов конечно или счетно, то $\bigcup_n A_n \leftrightarrow \bigcup_n \tilde{A}_n$.*

(C_4) *Если $A \leftrightarrow \tilde{A}$, то $P(A) = \tilde{P}(\tilde{A})$.*

(C_5) *Если $A \leftrightarrow \tilde{A}$, то $T^{-1}A \leftrightarrow \tilde{T}^{-1}\tilde{A}$.*

Это соответствие есть взаимно однозначное соответствие между классами эквивалентных множеств (в силу (C_1) и (C_2)). Оно сохраняет дополнения множеств и конечные или счетные их объединения (C_3), сохраняет меру (C_4) и замкнуто относительно операции обратного отображения (C_5).

Предположим теперь, что между полями $\mathcal{F}$ и $\tilde{\mathcal{F}}$ a priori существует некоторое многозначное соответствие, удовлетворяющее этим пяти условиям. Тогда говорят, что T и $\tilde{T}$ *сопряжены* [1]). Сопряженность, очевидно, является отношением эквивалентности.

Соответствие, определяющее сопряженность, автоматически обладает еще рядом свойств, кроме свойств (C_1) — (C_5). Из свойства (C_3) следует, что это соответствие сохраняет все конечные и счетные теоретико-множественные операции (например, разности, симметрические разности, конечные и счетные пересечения). Далее, пусть $T^{-1}A \leftrightarrow \tilde{T}^{-1}\tilde{A}$; в силу ($C_1$) существует такое множество $\tilde{B}$, что $A \leftrightarrow \tilde{B}$; из ($C_5$) следует $T^{-1}A \leftrightarrow \tilde{T}^{-1}\tilde{B}$, так что в силу ($C_2$) $\tilde{T}^{-1}\tilde{A} \sim \tilde{T}^{-1}\tilde{B}$ и, следовательно [2]), $\tilde{A} \sim \tilde{B}$, а тогда $A \leftrightarrow \tilde{A}$. Таким образом,

(C_6) *если* $T^{-1}A \leftrightarrow \tilde{T}^{-1}\tilde{A}$, *то* $A \leftrightarrow \tilde{A}$.

Если преобразования T и $\tilde{T}$ обратимы, то из свойств (C_5) и (C_6) следует, что $A \leftrightarrow \tilde{A}$ в том и только том случае, когда $TA \leftrightarrow \tilde{T}\tilde{A}$.

Мы видим, что если T и $\tilde{T}$ изоморфны, то они сопряжены. Пример 5.5 показывает, что обратное предложение, вообще говоря, неверно. Мы докажем, что такое обращение справедливо при дополнительных предположениях, что Ω и $\tilde{\Omega}$ — сепарабельные полные метрические пространства, а $\mathcal{F}$ и $\tilde{\mathcal{F}}$ являются σ-полями борелевских множеств этих пространств (т. е. $\mathcal{F}$ и $\tilde{\mathcal{F}}$ — это σ-поля, порожденные открытыми множествами, или, что то же самое в силу предположения сепарабельности, $\mathcal{F}$ и $\tilde{\mathcal{F}}$ — это σ-поля, порожденные шарами).

---

[1]) Такая формулировка устраняет необходимость введения алгебр с мерой. В терминах алгебр с мерой $A \sim B$ в том и только том случае, если $A + B$ принадлежит идеалу $\mathcal{N}$ нульмерных множеств и вышеупомянутое соответствие существует в том и только том случае, если фактор-алгебры $\mathcal{F}/\mathcal{N}$ и $\tilde{\mathcal{F}}/\tilde{\mathcal{N}}$ и преобразования на них абстрактно тождественны.

[2]) Заметим, что $A \sim B$ в том и только том случае, когда $T^{-1}A \sim T^{-1}B$ (аналогично для множеств пространства $\tilde{\Omega}$).

Все конкретные пространства $(\Omega, \mathcal{F})$, встречавшиеся нам в различных примерах, можно было бы определить в топологических терминах, но мы этого не делали. Если Ω — окружность, т. е. пространство, на котором определены вращения (пример 1.5), то оно имеет естественную метрику, относительно которой является сепарабельным полным метрическим пространством, а именно расстояние между двумя точками измеряется длиной кратчайшей дуги, их соединяющей; так как шарами здесь являются дуги, то соответствующее σ-поле $\mathcal{F}$ состоит из борелевских множеств в этой метрике.

Некоторые преобразования (преобразование, связанное с непрерывными дробями, и различные преобразования $T\omega = r\omega \pmod 1$) были определены на полуоткрытом единичном интервале $\Omega = [0, 1)$. В этом случае σ-поле состоит из обычных линейных борелевских подмножеств полуинтервала $[0, 1)$; это — σ-поле, порожденное открытыми подмножествами полуинтервала $[0, 1)$ с евклидовой метрикой. Относительно этой метрики пространство Ω сепарабельно, но неполно; так как полуинтервал $[0, 1)$ гомеоморфен лучу $[0, 1)$ в евклидовой метрике, то Ω может быть вновь метризовано так, чтобы быть не только сепарабельным, но и полным.

Мы рассматривали также преобразования на конечном пространстве Ω с классом всех его подмножеств в качестве поля $\mathcal{F}$. Здесь нам требовалась только дискретная метрика в Ω, скажем, такая, при которой расстояние между различными точками равно 1. (Ср. с пространством $\tilde{\Omega}$ примера 5.5, где $\tilde{\mathcal{F}}$ слишком мало.)

В пространстве $\Omega = \ldots \times \rho \times \rho \times \rho \times \ldots$ двустороннего сдвига σ-поле $\mathcal{F}$ порождено тонкими цилиндрами

$$\{\omega : x_l(\omega) = i_l,\ u \leqslant l \leqslant v\}. \tag{5.17}$$

Если расстояние между двумя точками ω и ω' определяется как

$$\sum_{n=-\infty}^{\infty} \frac{\delta(x_n(\omega), x_n(\omega'))}{2^{|n|}},$$

где $\delta(i, j)$ равняется 1 или 0 в зависимости от того, совпадают элементы i и j пространства состояний ρ или нет, то Ω является сепарабельным полным метрическим пространством. Если точка ω_0 принадлежит цилиндру (5.17) и $0 < \varepsilon < (1/2)^{|u|+|v|}$, то открытый шар радиуса ε с центром в точке ω_0 содержится в цилиндре (5.17). С другой стороны,

если $-u=v>0$, $(1/2)^{v-2}<\varepsilon$ и $i_l=x_l(\omega_0)$ для $|l|\leqslant v$, то точка ω_0 принадлежит цилиндру (5.17), который в свою очередь содержится в открытом шаре радиуса ε с центром в ω_0. Таким образом, тонкие цилиндры образуют базис в этой топологии, и в ней поле $\mathscr{F}$ совпадает с классом борелевских множеств. Подобным образом можно рассматривать и одностороннее прямое произведение $\rho\times\rho\times\ldots$.

Можно также показать[1]), что топологическое произведение бесконечной в обе стороны последовательности экземпляров любого сепарабельного полного метрического пространства может быть метризовано так, что оно станет сепарабельным полным пространством (пример 1.7).

Следующая теорема оправдывает выбор изоморфизма в качестве нашего понятия одинаковости сохраняющих меру преобразований.

Теорема 5.1. *Пусть T и $\tilde{T}$ — сохраняющие меру преобразования на $(\Omega, \mathscr{F}, P)$ и $(\tilde{\Omega}, \tilde{\mathscr{F}}, \tilde{P})$ соответственно, где Ω и $\tilde{\Omega}$ — сепарабельные полные метрические пространства, а $\mathscr{F}$ и $\tilde{\mathscr{F}}$ — это σ-поля борелевских множеств этих двух пространств. Если T и $\tilde{T}$ сопряжены, то они изоморфны.*

Для доказательства нам понадобятся три леммы, устанавливающие связь меры с топологией.

Лемма 1. *Для любого множества A из поля $\mathscr{F}$ и любого $\varepsilon>0$ существуют замкнутое множество F и открытое множество G, такие, что $F\subset A\subset G$ и $P(G-F)<\varepsilon$.*

Доказательство. Пусть $\mathscr{G}$ — класс множеств поля $\mathscr{F}$, обладающих указанным свойством. Достаточно показать, что $\mathscr{G}$ есть σ-поле, содержащее открытые множества. Очевидно, что класс $\mathscr{G}$ замкнут относительно взятия дополнения. Для заданных множеств A_n из $\mathscr{G}$ выберем замкнутые множества F_n и открытые множества G_n, такие, что $F_n\subset A_n\subset G_n$ и $P(G_n-F_n)<\varepsilon/2^{n+1}$. Если $F=\bigcup_n F_n$ и $G=\bigcup_n G_n$, то $F\subset A=\bigcup_n A_n\subset G$ и $P(G-F)<\varepsilon/2$. Множество G открыто; множество F не обязательно замкнуто, но его можно заменить конечным, следовательно, замкнутым, подобъединением $F_0=\bigcup_{n\leqslant n_0} F_n$, таким, что $P(F-F_0)<\varepsilon/2$. Отсюда следует, что

---

[1]) См., например, Данфорд и Шварц [1, стр. 44].

$A \in \mathscr{G}$. Итак, класс $\mathscr{G}$ есть σ-поле. Так как $\mathscr{G}$ замкнуто относительно взятия дополнения, то доказательство будет завершено, если мы покажем, что $\mathscr{G}$ содержит все замкнутые множества A. Для такого A множество G_k, состоящее из всех открытых шаров радиуса $1/k$ с центрами в точках из A, открыто и убывает к A; следовательно, мы можем взять A в качестве F и одно из множеств G_k в качестве G.

Лемма 2. *Если $A \in \mathscr{F}$ и $\varepsilon > 0$, то существует конечная или счетная совокупность попарно непересекающихся множеств $F_1, F_2, \ldots$, таких, что каждое F_n — замкнутое множество с диаметром, меньшим ε, $\bigcup\limits_n F_n \subset A$ и*

$$P\left(A - \bigcup_n F_n\right) = 0.$$

Доказательство. По лемме 1 A содержит такое замкнутое множество F_1, что $P(A - F_1) < 1$; $A - F_1$ содержит такое замкнутое множество F_2, что $P((A - F_1) - F_2) < 1/2$; $A - F_1 - F_2$ содержит такое замкнутое множество F_3, что

$$P((A - F_1 - F_2) - F_3) < \frac{1}{3},$$

и т. д. Построенные таким образом непересекающиеся замкнутые множества удовлетворяют соотношениям $\bigcup\limits_n F_n \subset A$ и и $P\left(A - \bigcup\limits_n F_n\right) = 0$.

Далее, так как пространство Ω по предположению сепарабельно, то A является объединением счетного числа непересекающихся измеримых множеств A_k, диаметры которых меньше ε. Аналогично строим для каждого A_k последовательность множеств $\{F_{kn}\}$. Все эти множества в совокупности удовлетворяют требованиям леммы.

Пусть $\mathscr{G}$ и $\mathscr{H}$ — два семейства попарно непересекающихся множеств. Мы говорим, что $\mathscr{H}$ измельчает $\mathscr{G}$, если каждый элемент семейства $\mathscr{H}$ является подмножеством некоторого элемента семейства $\mathscr{G}$. Из этого следует, что каждый элемент семейства $\mathscr{H}$ содержится ровно в одном элементе семейства $\mathscr{G}$, что если $H \in \mathscr{H}$, $G \in \mathscr{G}$ и $H \cap G \neq 0$, то $H \subset G$, и что объединение всех множеств из $\mathscr{H}$ содержится в объединении всех множеств из $\mathscr{G}$. Такие семейства являются разбиениями объединений своих элементов.

Лемма 3. *Пусть $\mathscr{C}_1, \mathscr{C}_2, \ldots$ — разбиения пространства Ω на такие конечные или счетные совокупности множеств из*

поля $\mathscr{F}$, что каждый элемент разбиения $\mathscr{C}_n$ имеет диаметр, меньший ε_n, где $\varepsilon_n \to 0$. Тогда $\bigcup\limits_n \mathscr{C}_n$ порождает $\mathscr{F}$. Пусть $\mathscr{D}_n$ состоит из конечных и счетных объединений элементов разбиения $\mathscr{C}_n$; если $\mathscr{C}_{n+1}$ измельчает $\mathscr{C}_n$ для каждого n, то $\mathscr{D}_1 \subset \mathscr{D}_2 \subset \ldots$ и $\bigcup\limits_n \mathscr{D}_n$ есть поле, порождающее $\mathscr{F}$.

Доказательство. Для любого заданного замкнутого множества F обозначим через A_n объединение тех элементов разбиения $\mathscr{C}_n$, которые пересекаются с F. Тогда $F \subset A_n$ и любая точка из A_n удалена не больше, чем на ε_n от некоторой точки множества F. Так как F замкнуто, отсюда следует, что $F = \bigcap\limits_n A_n$. Поэтому σ-поле, порожденное $\bigcup\limits_n \mathscr{C}_n$, содержит все замкнутые множества и, следовательно, совпадает с $\mathscr{F}$. Остальное тривиально.

Эти три леммы, разумеется, справедливы и по отношению к пространству $(\tilde{\Omega}, \tilde{\mathscr{F}}, \tilde{P})$.

Предположим теперь, что T и $\tilde{T}$ сопряжены относительно некоторого соответствия, удовлетворяющего условиям $(C_1)-(C_5)$. Для того чтобы доказать, что T и $\tilde{T}$ изоморфны, построим сначала две последовательности

$$\mathscr{C}_1, \mathscr{C}_2, \ldots$$

$$\tilde{\mathscr{C}}_1, \tilde{\mathscr{C}}_2, \ldots$$

классов со следующими свойствами.

(π_1) Класс $\mathscr{C}_n\,[\tilde{\mathscr{C}}_n]$ есть (непустое) конечное или счетное семейство попарно непересекающихся множеств из $\mathscr{F}\,[\tilde{\mathscr{F}}]$. Все элементы класса $\mathscr{C}_n\,[\tilde{\mathscr{C}}_n]$ имеют положительную P-меру [$\tilde{P}$-меру] и, следовательно, не пусты.

(π_2) Семейство $\mathscr{C}_{n+1}\,[\tilde{\mathscr{C}}_{n+1}]$ измельчает семейства $\mathscr{C}_n\,[\tilde{\mathscr{C}}_n]$ и $T^{-1}\mathscr{C}_n\,[\tilde{T}^{-1}\tilde{\mathscr{C}}_n]$.

(π_3) Для нечетных n [для четных n] множества из $\mathscr{C}_n\,[\tilde{\mathscr{C}}_n]$ замкнуты и имеют диаметры, меньшие $1/n$.

(π_4) Существует взаимно однозначное отображение $\mathscr{C}_n$ на $\tilde{\mathscr{C}}_n$, обладающее тем свойством, что если C — любой элемент класса $\mathscr{C}_n$ и $\tilde{C}$ — его образ в $\tilde{\mathscr{C}}_n$ относительно этого отображения, то $C \leftrightarrow \tilde{C}$.

(π_5) Если $\Omega_n\,[\tilde{\Omega}_n]$ — объединение всех элементов класса $\mathscr{C}_n\,[\tilde{\mathscr{C}}_n]$, то $P(\Omega_n) = 1$ $[\tilde{P}(\tilde{\Omega}_n) = 1]$.

Построим сначала $\mathscr{C}_1$ и $\tilde{\mathscr{C}}_1$. В силу леммы 2 существует конечное или счетное семейство непересекающихся замкнутых множеств, объединение которых Ω_1 таково, что $P(\Omega_1)=1$, причем элементы этого семейства имеют диаметр, меньший 1. Пусть класс $\mathscr{C}_1$ состоит из тех элементов этого семейства, которые имеют положительную меру.

Если $C_1, C_2, \ldots$ — элементы класса $\mathscr{C}_1$, то, согласно условию (C_1), существуют такие множества $\tilde{C}_1, \tilde{C}_2, \ldots$ из $\tilde{\mathscr{F}}$, что $C_u \longleftrightarrow \tilde{C}_u$. Семейство множеств $\tilde{C}_u - \bigcup\limits_{u' \neq u} \tilde{C}_{u'}$ удовлетворяет всем требованиям для $\tilde{\mathscr{C}}_1$.

Предположим, что построены частичные последовательности

$$\mathscr{C}_1, \mathscr{C}_2, \ldots, \mathscr{C}_n,$$

$$\tilde{\mathscr{C}}_1, \tilde{\mathscr{C}}_2, \ldots, \tilde{\mathscr{C}}_n.$$

Покажем, как построить классы $\mathscr{C}_{n+1}$ и $\tilde{\mathscr{C}}_{n+1}$. Возьмем сначала четное n (так что $n+1$ нечетно).

Пусть $C_1, C_2, \ldots$ — элементы класса $\mathscr{C}_n$, $\tilde{C}_1, \tilde{C}_2, \ldots$ — элементы класса $\tilde{\mathscr{C}}_n$, упорядоченные таким образом, что $C_u \longleftrightarrow \tilde{C}_u$. В силу леммы 2 каждое множество $C_{uv} = C_u \cap T^{-1}C_v$ содержит попарно непересекающиеся множества $C_{uv1}, C_{uv2}, \ldots$, замкнутые, имеющие диаметр, меньший $1/(n+1)$, и такие, что $P\left(C_{uv} - \bigcup\limits_w C_{uvw}\right) = 0$. Обозначим через $\tilde{D}_{uvw}$ множества из $\tilde{\mathscr{F}}$ такие, что $C_{uvw} \longleftrightarrow \tilde{D}_{uvw}$; положим $\tilde{E}_{uvw} = \tilde{D}_{uvw} \cap \tilde{C}_u \cap \cap \tilde{T}^{-1}\tilde{C}_v$, и пусть $\tilde{C}_{uvw} = \tilde{E}_{uvw} - \bigcup\limits_{w' \neq w} \tilde{E}_{uvw'}$. Из свойств ($C_1$)—($C_5$) соответствия, определяющего сопряженность, следует, что семейство $\mathscr{C}_{n+1}$, состоящее из множеств C_{uvw}, и семейство $\tilde{\mathscr{C}}_{n+1}$, состоящее из множеств $\tilde{C}_{uvw}$, обладают всеми требуемыми свойствами, за исключением того, что некоторые множества из этих семейств могут иметь меру 0. Выкинем все такие множества.

Если n нечетно (так что $n+1$ четно), мы просто изменим порядок построения и будем строить сначала класс $\tilde{\mathscr{C}}_{n+1}$, применяя лемму 2 к пространству $\tilde{\Omega}$, а затем с помощью отношения сопряженности класс $\mathscr{C}_{n+1}$.

Это завершает построение последовательностей $\mathscr{C}_1, \mathscr{C}_2, \ldots$ и $\tilde{\mathscr{C}}_1, \tilde{\mathscr{C}}_2, \ldots$. Если $\Omega_0 = \bigcap\limits_{n=1}^{\infty} \Omega_n$ и $\tilde{\Omega}_0 = \bigcap\limits_{n=1}^{\infty} \tilde{\Omega}_n$, то $P(\Omega_0) = 1$ и

$P(\tilde{\Omega}_0) = 1$. Мы определим отображение φ множества Ω_0 на $\tilde{\Omega}_0$, которое приведет к изоморфизму между преобразованиями T и $\tilde{T}$.

Будем называть *цепочкой* в Ω последовательность множеств $\{C_n\}$ таких, что $C_1 \supset C_2 \supset \ldots$ и $C_n \in \mathscr{C}_n$. Из того, что C_n непусто, и из свойства (π_3) следует, что пересечение множеств этой цепочки состоит из единственной точки, лежащей в Ω_0. И обратно, любая точка из Ω_0 определяет, таким образом, ровно одну цепочку. Все это справедливо и для $\tilde{\Omega}$.

Заменим класс $\mathscr{C}_n$ семейством $\{C \cap \Omega_0 : C \in \mathscr{C}_n\} = \mathscr{C}'_n$ и класс $\tilde{\mathscr{C}}_n$ семейством $\{\tilde{C} \cap \tilde{\Omega}_0 : \tilde{C} \in \tilde{\mathscr{C}}_n\} = \tilde{\mathscr{C}}'_n$. Тогда цепочки все еще пересекаются в единственной точке из Ω_0 или из $\tilde{\Omega}_0$ и условия $(\pi_1)-(\pi_5)$ все еще сохраняются, за исключением того, что от класса $\mathscr{C}_n$ при нечетном n и от класса $\tilde{\mathscr{C}}_n$ при четном n теперь не требуется, чтобы они состояли из замкнутых множеств (это больше не имеет значения). Семейство $\mathscr{C}'_n\,[\tilde{\mathscr{C}}'_n]$ является теперь разбиением множества $\Omega_0\,[\tilde{\Omega}_0]$.

Определим теперь отображение φ. Любая точка ω из Ω_0 определяет единственную цепочку $\{C_n\}$ в пространстве Ω, которая ввиду (π_4) соответствует единственной цепочке $\{\tilde{C}_n\}$ в $\tilde{\Omega}$ ($C_n \longleftrightarrow \tilde{C}_n$ для всех n), и эта вторая цепочка определяет единственную точку $\varphi\omega$ из $\tilde{\Omega}_0$. Так как все эти шаги можно проделать в обратном порядке, то отображение φ взаимно однозначно и область его значений совпадает со всем $\tilde{\Omega}_0$. Таким образом, условие (I_1) изоморфизма удовлетворяется.

Обратимся к условию (I_2). Теперь Ω_0 есть метрическое пространство с метрикой, унаследованной от Ω, и σ-полем $\mathscr{F}_0$ борелевских множеств из Ω_0, состоящим из тех подмножеств пространства Ω_0, которые лежат в $\mathscr{F}$. Пусть $\mathscr{G}_0$ — класс множеств A из $\mathscr{F}_0$, для которых $\varphi A \in \tilde{\mathscr{F}}$ и $P(A) = \tilde{P}(\varphi A)$. Если $C \in \mathscr{C}_n$, то φC есть то множество из $\tilde{\mathscr{C}}'_n$, которое соответствует множеству C по условию (π_4). Поэтому класс $\mathscr{G}_0$ содержит $\mathscr{C}_n$, а следовательно, и конечные и счетные объединения множеств, принадлежащих $\mathscr{C}_n$. Так как класс $\mathscr{G}_0$, очевидно, монотонный, то из леммы 3 следует, что $\mathscr{G}_0 = \mathscr{F}_0$. Эти и аналогичные результаты, где Ω и $\tilde{\Omega}$ меняются местами, показывают, что если $A \subset \Omega_0$, то $A \in \mathscr{F}$ в том и только том случае, когда $\varphi A \in \tilde{\mathscr{F}}$, и что при этом $P(A) = P(\varphi A)$. Поэтому условие (I_2) удовлетворяется.

Покажем, что если $\{C_n\}$ и $\{D_n\}$ — цепочки, определенные соответственно точками ω и ω' из Ω_0, то $\omega' = T\omega$ в том и только том случае, если $C_n \subset T^{-1}D_{n-1}$ для всех $n > 1$. (Аналогичное предложение верно для $\tilde{\Omega}_0$.) Если $C_n \subset T^{-1}D_{n-1}$ при $n > 1$, то обе точки $T\omega$ и ω' лежат в D_{n-1} при $n > 1$ и должны, следовательно, совпадать. С другой стороны, если $\omega' = T\omega$, то элемент C_n разбиения $\mathscr{C}_n$ и элемент $T^{-1}D_{n-1}$ разбиения $T^{-1}\mathscr{C}_{n-1}$ имеют общую точку ω. Так как первое из этих разбиений измельчает второе, то $C_n \subset T^{-1}D_{n-1}$.

Предположим теперь, что $\omega \in \Omega_0$ и $T\omega \in \Omega_0$, так что $\varphi\omega \in \tilde{\Omega}_0$ и $\varphi T\omega \in \tilde{\Omega}_0$. Пусть $\{C_n\}$, $\{D_n\}$; $\{\tilde{C}_n\}$ и $\{\tilde{D}_n\}$ — цепочки, определяемые точками ω, $T\omega$, $\varphi\omega$, $\varphi T\omega$ соответственно. Для того чтобы доказать, что $\tilde{T}\varphi\omega = \varphi T\omega$, достаточно проверить, что $\tilde{C}_n \subset \tilde{T}^{-1}\tilde{D}_{n-1}$. Но так как ω переходит в $T\omega$, то $C_n \subset T^{-1}D_{n-1}$. Далее, $C_n \leftrightarrow \tilde{C}_n$ и $D_{n-1} \leftrightarrow \tilde{D}_{n-1}$, так что $C_n = C_n \cap T^{-1}D_{n-1} \leftrightarrow \tilde{C}_n \cap \tilde{T}^{-1}\tilde{D}_{n-1}$. Поэтому $\tilde{C}_n \cap \tilde{T}^{-1}\tilde{D}_{n-1}$ имеет положительную меру и, значит, непусто; отсюда следует, что $\tilde{C}_n \subset \tilde{T}^{-1}\tilde{D}_{n-1}$.

Мы показали, что тройка $(\Omega_0, \tilde{\Omega}_0, \varphi)$ удовлетворяет условиям (I_1) и (I_2) определения изоморфизма и что она удовлетворяет условию (I_3), ослабленному в соответствии с замечанием 2, следующим за определением. Поэтому преобразования T и $\tilde{T}$ изоморфны.

Изоморфизм и спектральная эквивалентность *

Хотя эта книга лишь слегка касается изометрических операторов, порожденных сохраняющими меру преобразованиями, мы все же должны указать на различия между изоморфизмом и спектральной эквивалентностью.

Если T — сохраняющее меру преобразование на пространстве $(\Omega, \mathscr{F}, P)$, то равенство $(Uf)(\omega) = f(T\omega)$ определяет некоторое преобразование гильбертова пространства $L^2(\Omega)$ интегрируемых с квадратом функций на $(\Omega, \mathscr{F}, P)$. Как было показано в § 2, оператор U (там он обозначался $\tilde{T}$) изометрический. Покажем, что если T изоморфно некоторому другому преобразованию $\tilde{T}$, определенному на пространстве $(\tilde{\Omega}, \tilde{\mathscr{F}}, \tilde{P})$, то изометрический оператор U абстрактно совпадает с изометрическим оператором $\tilde{U}$, порожденным на пространстве $L^2(\tilde{\Omega})$ преобразованием $\tilde{T}$.

Пусть T и $\tilde{T}$ изоморфны относительно тройки $(\Omega_0, \tilde{\Omega}_0, \varphi)$. Для функции $\tilde{f}$ из пространства $L^2(\tilde{\Omega})$ обозначим через $V\tilde{f}$

функцию на Ω, принимающую значение $f(\varphi\omega)$ в точке ω. ($V\tilde{f}$ определена только на Ω_0; продолжим ее любым образом на Ω — элементы пространства $L^2(\Omega)$ определены с точностью до множеств меры 0.) Очевидно, что V является линейным отображением из $L^2(\tilde{\Omega})$ в $L^2(\Omega)$. Из свойств изоморфизма (I_1) и (I_2) следует, что V — взаимно однозначное отображение, сохраняющее длину и скалярное произведение, и область его значений совпадает со всем $L^2(\Omega)$. Наконец, из свойства (I_3) следует, что $V\tilde{U}=UV$, или $\tilde{U}=V^{-1}UV$.

Говорят, что операторы в гильбертовом пространстве, связанные посредством такого отображения V, имеют идентичную спектральную структуру, или спектрально эквивалентны; это — понятие одинаковости применительно к операторам. Мы показали, что спектральная структура порожденного изометрического оператора инвариантна по отношению к изоморфизму. (Она является инвариантом и по отношению к сопряженности.) Если операторы U и $\tilde{U}$ спектрально эквивалентны, то можно говорить, что сами преобразования T и $\tilde{T}$ спектрально эквивалентны.

Функция f, принадлежащая пространству $L^2(\Omega)$, является инвариантной ($f(T\omega)=f(\omega)$ п. в) в том и только том случае, если $Uf=f$. Далее, преобразование T эргодично в том и только том случае, если все инвариантные функции являются почти всюду константами. Эти функции как элементы пространства $L^2(\Omega)$ получаются друг из друга умножением на скаляры, поэтому T эргодично в том и только том случае, когда подпространство решений уравнения $Uf=f$ одномерно.

Если это подпространство одномерно, то то же верно для любого оператора, спектрально эквивалентного U. Поэтому если известна спектральная структура оператора U, то известно, в частности, является ли преобразование T эргодическим. Таким образом, спектральная структура порожденного изометрического оператора — более точный инвариант, чем эргодичность: если T и $\tilde{T}$ спектрально эквивалентны, то они либо оба эргодичны, либо оба неэргодичны. Можно также показать (см. Халмош [3]), что спектральная структура — инвариант более точный по сравнению с перемешиванием.

Покажем теперь, что сдвиги Бернулли порождают изометрические операторы, имеющие совершенно одинаковую спектральную структуру. При обсуждении энтропии уже указывалось, что из теоремы Колмогорова (которую мы докажем в § 7) следует, что существуют сдвиги Бер-

нулли, имеющие различные энтропии и потому неизоморфные. Таким образом, энтропия — инвариант, дающий возможность различать некоторые преобразования, которые не различаются с помощью спектральной структуры, и, разумеется, бесполезно пытаться различить их с помощью таких инвариантов, как эргодичность и перемешивание, ибо они еще менее точны, чем спектральная структура. (С другой стороны, имеются спектрально различные преобразования с одинаковой энтропией, например циклические перестановки k точек равной массы для различных k.)

Пусть T — сдвиг Бернулли. Возьмем пространство состояний $\rho=\{1, 2, \ldots, r\}$ и вероятности $p_1, p_2, \ldots, p_r$ и предположим, что все p_i положительны. Пусть векторы $\xi_i = (\xi_i(1), \ldots, \xi_i(r))$, $i=1, 2, \ldots, r$, таковы, что

$$\sum_k p_k \xi_i(k)\, \xi_j(k) = \delta_{ij}, \qquad i, j = 1, 2, \ldots, r, \tag{5.18}$$

и

$$\xi_1 = (1, 1, \ldots, 1). \tag{5.19}$$

Иными словами, векторы $\left(\xi_i(1)\sqrt{p_1}, \ldots, \xi_i(r)\sqrt{p_r}\right)$ образуют ортонормированный базис в r-мерном пространстве.

Рассмотрим бесконечную в обе стороны последовательность $u=(\ldots, u_{-1}, u_0, u_1, \ldots)$ элементов пространства $\{1, 2, \ldots, r\}$. Пусть $\mathscr{U}$ — совокупность таких последовательностей, у которых лишь конечное число координат u_n отлично от 1. Определим для последовательности u из $\mathscr{U}$ функцию g_u на пространстве Ω формулой

$$g_u(\omega) = \prod_{n=-\infty}^{\infty} \xi_{u_n}(x_n(\omega)), \tag{5.20}$$

где x_n — координатные переменные. Так как лишь конечное число u_n отлично от 1, то произведение (5.20) на самом деле конечно (см. 5.19).

Функции g_u являются элементами пространства $L^2(\Omega)$. Так как переменные x_n независимы относительно меры P, то в силу (5.18) имеем

$$(g_u, g_v) = \prod_{n=-\infty}^{\infty} \int \xi_{u_n}(x_n(\omega))\, \xi_{v_n}(x_n(\omega))\, P(d\omega) = \prod_{n=-\infty}^{\infty} \delta_{u_n v_n}, \quad u, v \in \mathscr{U}.$$

Итак, функции g_u ортонормированы.

Любую функцию из $L^2(\Omega)$ можно аппроксимировать в смысле L^2 простыми функциями $\sum_{i=1}^{k} \alpha_i I_{A_i}$. Так как множества A_i можно аппроксимировать множествами, зависящими только от конечного числа координат (цилиндрами), то функции, зависящие лишь от конечного числа координат, порождают $L^2(\Omega)$. Но функцию, зависящую, скажем, только от n координат, можно идентифицировать с точкой r^n-мерного пространства. Так как ξ_i порождают r-мерное пространство, то такая функция является конечной линейной комбинацией функций g_u. Поэтому $\{g_u : u \in \mathscr{U}\}$ — ортонормированный базис в $L^2(\Omega)$.

Определим отображение $\theta : \mathscr{U} \to \mathscr{U}$ соотношением $(\theta u)_n = u_{n-1}$. Так как

$$(Ug_u)(\omega) = g_u(T\omega) = \prod_{n=-\infty}^{\infty} \xi_{u_{n-1}}(x_n(\omega)),$$

то

$$Ug_u = g_{\theta u}.$$

Пусть теперь $u^{(0)}$, $u^{(1)}, \ldots$ — такая последовательность элементов из $\mathscr{U}$, что $u^{(0)}$ состоит целиком из единиц, $u^{(i)} \neq \theta^n u^{(j)}$ при $i \neq j$ для всех $n = 0, \pm 1, \ldots$ ($u^{(i)}$ не являются сдвигами друг друга) и любой элемент u из $\mathscr{U}$ можно представить в виде $u = \theta^n u^{(i)}$ для некоторого $i = 0, 1, \ldots$ и некоторого $n = 0, \pm 1, \ldots$ (любой элемент u есть сдвиг некоторого $u^{(i)}$). Пусть $f_0 = g_{u^{(0)}}$ и $f_{i,n} = g_{\theta^n u^{(i)}}$ для $i = 1, 2, \ldots$ и $n = 0, \pm 1, \ldots$. Тогда система $\{g_u : u \in \mathscr{U}\}$ принимает вид

$$\begin{array}{c} f_0 \\ \ldots, f_{1,-1}, f_{1,0}, f_{1,1}, \ldots \\ \ldots, f_{2,-1}, f_{2,0}, f_{2,1}, \ldots \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \end{array}$$

Элементы этой системы образуют ортонормированный базис в $L^2(\Omega)$; $Uf_0 = f_0$ и $Uf_{i,n} = f_{i,n+1}$ для $i \geqslant 1$ и $n = 0, \pm 1, \ldots$. Такая спецификация спектральной структуры оператора U не зависит от r и от вероятностей $p_1, p_2, \ldots, p_r$.

Мы показали, что изоморфные сохраняющие меру преобразования спектрально эквивалентны. Только что полученный результат вместе с фактом существования неизоморфных сдвигов Бернулли показывает, что спектрально эквивалентные преобразования совсем не обязаны быть изо-

морфными (а также сопряженными). Иными словами, спектральная структура порожденного изометрического оператора не является полным инвариантом.

Замечание. По поводу примера 5.4 см. Харрис [1].

Понятие энтропии было введено в теории связи Шенноном [1]. Колмогоров [2] определил энтропию общего сохраняющего меру преобразования; важный вклад был сделан Синаем [1].

Теоремой 5.1 мы обязаны Нейману [2]. Последующую информацию и ссылки можно получить у Якобса [1, 2].

В книге Халмоша [3] дано широкое обсуждение порожденных изометрических операторов.

6. СВОЙСТВА ФУНКЦИЙ $H(\mathscr{A})$ И $h(\mathscr{A}, T)$

Полезным вспомогательным понятием является понятие энтропии (или условной энтропии) поля $\mathscr{A}$ при заданном поле $\mathscr{B}$, определяемой формулой

$$H(\mathscr{A} \mid \mathscr{B}) = \sum_B P(B) \sum_A \eta(P(A \mid B)) = -\sum_{A, B} P(A \cap B) \ln P(A \mid B), \qquad (6.1)$$

где суммирование производится по атомам полей $\mathscr{A}$ и $\mathscr{B}$, за исключением тех атомов B поля $\mathscr{B}$, которые имеют меру 0.

Тихе, выбирающая некоторую точку ω из пространства Ω в соответствии с вероятностной мерой P, сообщает экспериментатору, какому атому поля $\mathscr{B}$ эта точка принадлежит. Однако обычно у экспериментатора нет уверенности относительно того, в каком атоме поля $\mathscr{A}$ лежит точка ω, и степень его неуверенности измеряется величиной $\sum\limits_A \eta(P(A \mid B))$, где B — тот атом поля $\mathscr{B}$, который содержит ω. Таким образом, величина (6.1) измеряет среднее количество неопределенности относительно исхода эксперимента $\mathscr{A}$ при условии, что известен исход эксперимента $\mathscr{B}$. Или, иначе, величина $H(\mathscr{A} \mid \mathscr{B})$ измеряет количество дополнительной информации, получаемой экспериментатором при узнавании исхода $\mathscr{A}$ в предположении, что он уже знает исход $\mathscr{B}$.

Если обозначить символом 2 поле, состоящее из пространства Ω и пустого множества, то, очевидно, что $H(\mathscr{A} \mid 2) = H(\mathscr{A})$.

Свойства функций $H(\mathcal{A})$ и $H(\mathcal{A}\mid\mathcal{B})$

Выведем следующие пять пар соотношений, касающихся энтропии и условной энтропии для конечных полей. Основной является первая формула:

$$(\mathrm{A}_1)\qquad H(\mathcal{A}\vee\mathcal{B}\mid\mathcal{C})=H(\mathcal{A}\mid\mathcal{C})+H(B\mid\mathcal{A}\vee\mathcal{C}),$$

$$(\mathrm{A}_1')\qquad H(\mathcal{A}\vee\mathcal{B})=H(\mathcal{A})+H(\mathcal{B}\mid\mathcal{A}).$$

Формула (A_1') означает, что информация, содержащаяся в экспериментах $\mathcal{A}$ и $\mathcal{B}$ вместе, равна информации, содержащейся в $\mathcal{A}$, плюс информация, содержащаяся в $\mathcal{B}$ при условии, что результат эксперимента $\mathcal{A}$ известен. Аналогичная интерпретация имеет место для формулы (A_1).

Условная энтропия не убывает по своему первому аргументу:

$$(\mathrm{A}_2)\qquad H(\mathcal{A}\mid\mathcal{C})\leqslant H(\mathcal{B}\mid\mathcal{C}),\quad \text{если}\quad \mathcal{A}\subset\mathcal{B},$$

$$(\mathrm{A}_2')\qquad H(\mathcal{A})\leqslant H(\mathcal{B}),\quad \text{если}\quad \mathcal{A}\subset\mathcal{B}.$$

Это означает, что если $\mathcal{A}\subset\mathcal{B}$, то $\mathcal{B}$ измельчает $\mathcal{A}$; знание исхода эксперимента $\mathcal{B}$ влечет знание исхода эксперимента $\mathcal{A}$; следовательно, $\mathcal{B}$ более информативно.

Условная энтропия не возрастает по своему второму аргументу:

$$(\mathrm{A}_3)\qquad H(\mathcal{A}\mid\mathcal{C})\leqslant H(\mathcal{A}\mid\mathcal{B}),\quad \text{если}\quad \mathcal{C}\supset\mathcal{B},$$

$$(\mathrm{A}_3')\qquad H(\mathcal{A}\mid\mathcal{C})\leqslant H(\mathcal{A}).$$

Если знание исхода эксперимента $\mathcal{C}$ влечет знание исхода эксперимента $\mathcal{B}$, то остающаяся экспериментатору неопределенность относительно исхода эксперимента $\mathcal{A}$ должна быть меньше, если он знает исход эксперимента $\mathcal{C}$, чем если он знает только исход эксперимента $\mathcal{B}$.

Условная энтропия полуаддитивна по первому аргументу:

$$(\mathrm{A}_4)\qquad H(\mathcal{A}\vee\mathcal{B}\mid\mathcal{C})\leqslant H(\mathcal{A}\mid\mathcal{C})+H(\mathcal{B}\mid\mathcal{C}),$$

$$(\mathrm{A}_4')\qquad H(\mathcal{A}\vee\mathcal{B})\leqslant H(\mathcal{A})+H(\mathcal{B}).$$

Наконец,

$$(\mathrm{A}_5)\qquad H\left(T^{-1}\mathcal{A}\mid T^{-1}\mathcal{B}\right)=H(\mathcal{A}\mid\mathcal{B}),$$

$$(\mathrm{A}_5')\qquad H\left(T^{-1}\mathcal{A}\right)=H(\mathcal{A}).$$

Заметим, что для любого i формула (A_i') немедленно получается из (A_i), если заменить соответствующее поле полем 2.

Соотношение (A_1) аналогично формуле для условных вероятностей

$$P(A \cap B \mid C) = P(A \mid C) P(B \mid A \cap C).$$

В самом деле, (A_1) следует из этой формулы и того факта, что атомы полей $\mathcal{A} \vee \mathcal{B}$ и $\mathcal{A} \vee \mathcal{C}$ являются соответственно множествами вида $A \cap B$ и $A \cap C$, где A, B и C — атомы полей $\mathcal{A}$, $\mathcal{B}$ и $\mathcal{C}$:

$$\begin{aligned} H(\mathcal{A} \vee \mathcal{B} \mid \mathcal{C}) &= -\sum_{A, B, C} P(A \cap B \cap C) \ln P(A \cap B \mid C) = \\ &= -\sum_{A, B, C} P(A \cap B \cap C) \ln P(A \mid C) - \\ &\quad -\sum_{A, B, C} P(A \cap B \cap C) \ln P(B \mid A \cap C) = \\ &= -\sum_{A, C} P(A \cap C) \ln P(A \mid C) - \\ &\quad -\sum_{A, B, C} P(B \cap (A \cap C)) \ln P(B \mid A \cap C) = \\ &= H(\mathcal{A} \mid \mathcal{C}) + H(\mathcal{B} \mid \mathcal{A} \vee \mathcal{C}). \end{aligned}$$

Если $\mathcal{A} \subset \mathcal{B}$, то $\mathcal{A} \vee \mathcal{B} = \mathcal{B}$; следовательно, ($A_2$) вытекает из ($A_1$) и очевидной неотрицательности условной энтропии.

Для доказательства соотношения (A_3) заметим, что так как функция $\eta(t) = -t \ln t$ выпукла, то

$$\sum_C \eta(P(A \mid C)) P(C \mid B) \leqslant \eta\Big(\sum_C P(A \mid C) P(C \mid B)\Big).$$

Так как $\mathcal{B} \subset \mathcal{C}$, то B есть объединение атомов поля $\mathcal{C}$; следовательно,

$$\sum_C P(A \mid C) P(C \mid B) = \sum_{C \subset B} \frac{P(A \cap C)}{P(C)} \frac{P(C)}{P(B)} = P(A \mid B).$$

Поэтому

$$\sum_C \eta(P(A \mid C)) P(C \mid B) \leqslant \eta(P(A \mid B)).$$

Умножая это неравенство на $P(B)$ и суммируя его по атомам A и B, получаем (A_3).

Наконец, (A_4) следует непосредственно из (A_1) и (A_3). Соотношение (A_5) очевидно.

Если $\mathcal{A} = \mathcal{B}$, то $H(\mathcal{A} \mid \mathcal{B}) = 0$ ввиду (A_1'). Интуитивно ясно, что если $\mathcal{A}$ и $\mathcal{B}$ почти равны в том смысле, что каждый атом одного из них отличается от некоторого атома другого на множество малой меры, то величина $H(\mathcal{A} \mid \mathcal{B})$

должна быть малой. Эта идея лежит в основе доказательства следующего результата.

Теорема 6.1. *Предположим, что конечное поле $\mathcal{A}$ содержится в σ-поле, порожденном полем $\mathcal{F}_0$ (или, более общо, каждый атом поля $\mathcal{A}$ отличается на множество меры 0 от некоторого элемента σ-поля, порожденного полем $\mathcal{F}_0$). Тогда для любого положительного ε существует такое конечное подполе $\mathcal{B}$ поля $\mathcal{F}_0$, что $H(\mathcal{A}\,|\,\mathcal{B})<\varepsilon$.*

Доказательство. Не ограничивая общности, предположим, что все атомы $A_1, \ldots, A_r$ поля $\mathcal{A}$ имеют положительную меру. Так как функция $\eta(t)$ непрерывна и $\eta(0)=\eta(1)=0$, то существует такое число δ_0 $(0<\delta_0<1)$, что $\eta(t)<\varepsilon/r$, когда либо $0\leqslant t\leqslant\delta_0$, либо $1-\delta_0\leqslant t\leqslant 1$.

Если мы сможем указать в $\mathcal{F}_0$ конечное подполе $\mathcal{B}$, атомы которого $B_1, \ldots, B_r$ удовлетворяют условию

$$P(A_i\,|\,B_i)>1-\delta_0, \qquad i=1, \ldots, r,$$

то получим $P(A_j\,|\,B_i)<\delta_0$ для $j\neq i$ и, следовательно,

$$H(\mathcal{A}\,|\,\mathcal{B})=\sum_{ij}P(B_j)\,\eta(P(A_i\,|\,B_j))<\sum_{ij}P(B_j)\frac{\varepsilon}{r}=\varepsilon.$$

Если атомы B_i удовлетворяют условию

$$P(A_i+B_i)<\delta=\min_{1\leqslant i\leqslant r}\delta_0\frac{P(A_i)}{2}, \tag{6.2}$$

то, так как $P(A_i)\leqslant P(B_i)+\delta<P(B_i)+P(A_i)/2$, будем иметь $P(A_i)/2<P(B_i)$ и, следовательно, $P(B_i)-P(A_i\cap B_i)<\delta<\delta_0 P(B_i)$, или $P(A_i\,|\,B_i)>1-\delta_0$. Поэтому достаточно получить поле $\mathcal{B}$ с атомами, удовлетворяющими условию (6.2).

В силу предположения теоремы для каждого i существует такое множество B_i' из поля $\mathcal{F}_0$, что $P(A_i+B_i')<\lambda$, где λ будет выбрано несколько позже. Если $i\neq j$, то $P(B_i'\cap B_j')\leqslant P(B_i'+A_i)+P(A_j+B_j')<2\lambda$, так что $P(N)<r(r-1)\lambda$, где $N=\bigcup_{i\neq j}(B_i'\cap B_j')$. Определим $B_i=B_i'-N$ для $1\leqslant i<r$ и $B_r=\Omega-\bigcup_{i<r}B_i$; тогда множество B_i принадлежит полю $\mathcal{F}_0$ и $P(A_i+B_i)<\lambda+r(r-1)\lambda$ для $i<r$; следовательно, $P(A_r+B_r)<(r-1)(\lambda+r(r-1)\lambda)$. Если λ достаточно мало, то (6.2) выполняется.

Следующая теорема иным образом уточняет идею, что если атомы полей $\mathcal{A}$ и $\mathcal{B}$ близки, то величина $H(\mathcal{A}\,|\,\mathcal{B})$

мала. Теорема эта будет использована только в гл. 5, в теории кодирования.

Теорема 6.2. *Пусть число атомов $A_1, \ldots, A_r$ поля $\mathcal{A}$ равно числу атомов $B_1, \ldots, B_r$ поля $\mathcal{B}$. Тогда*

$$H(\mathcal{A} \mid \mathcal{B}) \leqslant \eta(d) + \eta(1-d) + d \ln(r-1),$$

где

$$d = \sum_j P(B_j) P(A_j^c \mid B_j) = \sum_{i \neq j} P(A_i \cap B_j) = \sum_i P(A_i) P(B_i^c \mid A_i).$$

Доказательство. Прежде всего можно записать

$$H(\mathcal{A} \mid \mathcal{B}) = \sum_j P(B_j)\left[\eta(P(A_j \mid B_j)) + \sum_{i \neq j} \eta(P(A_i \mid B_j))\right].$$

Для фиксированного j имеем [1])

$$\sum_{i \neq j} \eta(P(A_i \mid B_j)) = P(A_j^c \mid B_j) \sum_{i \neq j} \eta\left(\frac{P(A_i \mid B_j)}{P(A_j^c \mid B_j)}\right) + \\ + \eta(P(A_j^c \mid B_j)) \leqslant P(A_j^c \mid B_j) \ln(r-1) + \eta(P(A_j^c \mid B_j)).$$

Следовательно,

$$H(\mathcal{A} \mid \mathcal{B}) \leqslant \sum_j P(B_j)\left[\eta(P(A_j \mid B_j)) + \eta(P(A_j^c \mid B_j))\right] + \\ + \sum_j P(B_j) P(A_j^c \mid B_j) \ln(r-1).$$

Вторая сумма равна $d \ln(r-1)$. Если $\mathcal{E}$ — поле с атомами $E = \bigcup_{i \neq j} (A_i \cap B_j)$ и E^c, то первая сумма равна $H(\mathcal{E} \mid \mathcal{B})$. А так как

$$H(\mathcal{E} \mid \mathcal{B}) \leqslant H(\mathcal{E}) = \eta(P(E)) + \eta(P(E^c)) = \eta(d) + \eta(1-d),$$

то доказательство завершено.

Если $P(B_j^c \mid A_j) \leqslant \varepsilon$ для всех j, то $d \leqslant \varepsilon$; использование этого факта приводит к несколько более экономному доказательству теоремы 6.1.

---

[1]) Для неотрицательных чисел $t_1, \ldots, t_k$, дающих в сумме 1, справедливо неравенство $\sum_{i=1}^{k} \eta(t_i) \leqslant \ln k$.

Свойства функции $h(\mathcal{A}, T)$

Докажем сначала, что верхний предел в определении (5.13) функции $h(\mathcal{A}, T)$ может быть заменен обычным пределом. Действительно, (A_1') и (A_5') означают, что

$$H\left(\mathcal{A}\,\middle|\,\bigvee_{i=1}^{k} T^{-i}\mathcal{A}\right) = H\left(\bigvee_{i=0}^{k} T^{-i}\mathcal{A}\right) - H\left(\bigvee_{i=1}^{k} T^{-i}\mathcal{A}\right) = \\ = H\left(\bigvee_{i=0}^{k} T^{-i}\mathcal{A}\right) - H\left(\bigvee_{i=0}^{k-1} T^{-i}\mathcal{A}\right).$$

Суммируя по k, имеем

$$H\left(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{A}\right) = H(\mathcal{A}) + \sum_{k=1}^{n-1} H\left(\mathcal{A}\,\middle|\,\bigvee_{i=1}^{k} T^{-i}\mathcal{A}\right).$$

В силу (A_3) функция $H\left(\mathcal{A}\,\middle|\,\bigvee_{i=1}^{k} T^{-i}\mathcal{A}\right)$ не возрастает по k и, следовательно, имеет конечный предел. Беря среднее по Чезаро, получаем из предыдущего равенства

$$(B_1)\quad h(\mathcal{A}, T) = \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{A}\right) = \lim_{n\to\infty} H\left(\mathcal{A}\,\middle|\,\bigvee_{i=1}^{n} T^{-i}\mathcal{A}\right).$$

Кроме того, обе последовательности, пределы которых берутся в (B_1), невозрастающие. Аналогичное рассуждение приводит к соотношению

$$(B_1')\qquad h(\mathcal{A}, T) = \lim_{n\to\infty} H\left(T^{-n}\mathcal{A}\,\middle|\,\bigvee_{i=0}^{n-1} T^{-i}\mathcal{A}\right),$$

где, как и выше, стремление к пределу монотонно. Если T обратимо, то

$$(B_1'')\qquad h(\mathcal{A}, T) = \lim_{n\to\infty} H\left(\mathcal{A}\,\middle|\,\bigvee_{i=1}^{n} T^{i}\mathcal{A}\right),$$

что является результатом применения T^n к правой части равенства (B_1').

Соотношение (B_1') дает еще одну интерпретацию функции $h(\mathcal{A}, T)$. Количество неопределенности относительно исхода $(n+1)$-й реализации $T^{-n}\mathcal{A}$ эксперимента $\mathcal{A}$ при данных исходах первых n реализаций $\mathcal{A}, T^{-1}\mathcal{A}, \ldots, T^{-(n-1)}\mathcal{A}$ измеряется величиной $H\left(T^{-n}\mathcal{A}\,\middle|\,\bigvee_{i=0}^{n-1} T^{-i}\mathcal{A}\right)$. Эта величина, которая измеряет также количество информации, доставляе-

мой $(n+1)$-й реализацией в добавление к уже имеющейся от первых n реализаций, стремится к $h(\mathcal{A}, T)$ при $n\to\infty$.

Из соотношений (B_1) и (A_2') непосредственно следует

$$(B_2)\qquad h(\mathcal{A}, T)\leqslant h(\mathcal{B}, T),\quad \text{если}\quad \mathcal{A}\subset\mathcal{B}.$$

Если $u\leqslant v$ (и $u\geqslant 0$, если преобразование T необратимо), то

$$\bigvee_{i=0}^{n-1} T^{-i}\left\{\bigvee_{j=u}^{v} T^{-j}\mathcal{A}\right\}=T^{-u}\left\{\bigvee_{k=0}^{n+v-u-1} T^{-k}\mathcal{A}\right\},$$

так что в силу (A_5')

$$\frac{1}{n}H\left(\bigvee_{i=0}^{n-1} T^{-i}\left\{\bigvee_{j=u}^{v} T^{-j}\mathcal{A}\right\}\right)=\frac{n+v-u}{n}\,\frac{1}{n+v-u}H\left(\bigvee_{k=0}^{n+v-u-1} T^{-k}\mathcal{A}\right),$$

Из (B_1) следует

$$(B_3)\qquad h\left(\bigvee_{j=u}^{v} T^{-j}\mathcal{A},\ T\right)=h(\mathcal{A}, T),$$

Если $u=v=1$, то (B_3) превращается в

$$h(T^{-1}\mathcal{A}, T)=h(\mathcal{A}, T),$$

а при $u=0$

$$h\left(\bigvee_{j=0}^{v} T^{-j}\mathcal{A},\ T\right)=h(\mathcal{A}, T),$$

откуда видно, что информация, приходящаяся на эксперимент в последовательности экспериментов

$$\bigvee_{j=0}^{v} T^{-j}\mathcal{A},\quad \bigvee_{j=1}^{v+1} T^{-j}\mathcal{A},\quad \bigvee_{j=2}^{v+2} T^{-j}\mathcal{A},\ \ldots,$$

такая же, как в последовательности

$$\mathcal{A},\ T^{-1}\mathcal{A},\ T^{-2}\mathcal{A},\ \ldots.$$

Это естественно вследствие того, что соседние эксперименты первой из этих последовательностей перекрываются.

Если $k\geqslant 1$, то

$$\frac{1}{n}H\left(\bigvee_{i=0}^{n-1}(T^k)^{-i}\left\{\bigvee_{j=0}^{k-1} T^{-j}\mathcal{A}\right\}\right)=k\,\frac{1}{nk}H\left(\bigvee_{u=0}^{nk-1} T^{-u}\mathcal{A}\right),$$

так что опять в силу (B_1)

$$(B_4)\qquad h\left(\bigvee_{j=0}^{k-1} T^{-j}\mathcal{A},\ T^k\right)=kh(\mathcal{A}, T),\quad k\geqslant 1.$$

И, наконец, еще одно нужное нам свойство функции $h(\mathcal{A}, T)$, являющееся обобщением неравенства (B_2):

$$(B_5) \qquad h(\mathcal{A}, T) \leqslant h(\mathcal{B}, T) + H(\mathcal{A} \mid \mathcal{B}).$$

Для доказательства этого соотношения заметим, что в силу (A_2') и (A_1')

$$H\left(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{A}\right) \leqslant H\left(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{A} \vee \bigvee_{j=0}^{n-1} T^{-j}\mathcal{B}\right) = \\ = H\left(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{B}\right) + H\left(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{A} \middle| \bigvee_{j=0}^{n-1} T^{-j}\mathcal{B}\right).$$

Последовательно применяя (A_4), (A_3) и (A_5), имеем

$$H\left(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{A} \middle| \bigvee_{j=0}^{n-1} T^{-j}\mathcal{B}\right) \leqslant \sum_{i=0}^{n-1} H\left(T^{-i}\mathcal{A} \middle| \bigvee_{j=0}^{n-1} T^{-j}\mathcal{B}\right) \leqslant \\ \leqslant \sum_{i=0}^{n-1} H\left(T^{-i}\mathcal{A} \mid T^{-i}\mathcal{B}\right) = nH(\mathcal{A} \mid \mathcal{B}).$$

Комбинируя предыдущие неравенства, получаем

$$\frac{1}{n} H\left(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{A}\right) \leqslant \frac{1}{n} H\left(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{B}\right) + H(\mathcal{A} \mid \mathcal{B}),$$

откуда в силу (B_1) следует (B_5).

7. СВОЙСТВА ФУНКЦИИ $h(T)$

Теорема Колмогорова

Мы докажем теперь теорему Колмогорова, которая позволит нам во многих случаях вычислить энтропию

$$h(T) = \sup h(\mathcal{A}, T).$$

Теорема 7.1. *Если T обратимо и* $\bigvee_{n=-\infty}^{\infty} T^n\mathcal{A} = \mathcal{F}$ [1]), *то* $h(T) = h(\mathcal{A}, T)$.

Доказательство. Мы должны показать, что если $\mathcal{B}$ — любое конечное подполе поля $\mathcal{F}$, то

$$h(\mathcal{B}, T) \leqslant h(\mathcal{A}, T).$$

---

[1]) Разбиение на атомы поля $\mathcal{A}$ называется в этом случае *образующим разбиением* для T или просто *образующей*. — *Прим. ред.*

Пусть $\mathcal{A}_n = \bigvee_{k=-n}^{n} T^k\mathcal{A}$; из ($B_3$) следует

$$h(\mathcal{A}_n, T) = h(\mathcal{A}, T),$$

а из (B_5)

$$h(\mathcal{B}, T) \leqslant h(\mathcal{A}_n, T) + H(\mathcal{B} \mid \mathcal{A}_n) = h(\mathcal{A}, T) + H(\mathcal{B} \mid \mathcal{A}_n).$$

Поэтому достаточно показать, что

$$\lim_{n\to\infty} H(\mathcal{B} \mid \mathcal{A}_n) = 0.$$

Докажем это с помощью теоремы 6.1.

Если $\mathcal{F}_0 = \bigcup_n \mathcal{A}_n$, то $\mathcal{F}_0$ является (конечно аддитивным) полем, порождающим поле $\mathcal{F}$. В силу теоремы 6.1 для любого положительного ε существует такое конечное подполе $\mathcal{C}$ поля $\mathcal{F}_0$, что $H(\mathcal{B}\mid\mathcal{C}) < \varepsilon$. Далее, $\mathcal{C}$ принадлежит некоторому $\mathcal{A}_{n_0}$; если $n \geqslant n_0$, то в силу (A_3)

$$H(\mathcal{B} \mid \mathcal{A}_n) \leqslant H(\mathcal{B} \mid \mathcal{A}_{n_0}) \leqslant H(\mathcal{B} \mid \mathcal{C}) < \varepsilon,$$

что доказывает теорему.

Замена $\mathcal{A}_n$ на $\bigvee_{k=0}^{n} T^{-k}\mathcal{A}$ в этом доказательстве приводит к следующему результату, для которого не требуется обратимости T.

Следствие. *Если* $\bigvee_{n=0}^{\infty} T^{-n}\mathcal{A} = \mathcal{F}$, *то* $h(T) = h(\mathcal{A}, T)$.

Вычисление энтропии

Теорема 7.1 сделала законными вычисления энтропии в § 5. Проведем еще раз эти вычисления, но несколько иным способом. Пусть T — общий двусторонний сдвиг, соответствующий процессу $\{x_n\}$ с конечным пространством состояний ρ (пример 1.2), и пусть $\mathcal{A}$ — поле событий, наблюдавшихся в момент времени 0 (поле с атомами $\{x_0 = i\}$, $i \in \rho$). Так как $\bigvee_{n=-\infty}^{\infty} T^n\mathcal{A} = \mathcal{F}$, то из теоремы 7.1 следует $h(T) = h(\mathcal{A}, T)$. Вычислим $h(\mathcal{A}, T)$ по формуле

$$h(\mathcal{A}, T) = \lim_{n\to\infty} H\left(\mathcal{A} \,\Big|\, \bigvee_{k=1}^{n} T^k\mathcal{A}\right).$$

Произвольный атом поля $\bigvee_{k=1}^{n} T^k \mathcal{A}$ имеет вид $\{x_{-1} = i_{-1}, \ldots$ $\ldots, x_{-n} = i_{-n}\}$, и, следовательно,

$$H\left(\mathcal{A} \mid \bigvee_{k=1}^{n} T^k \mathcal{A}\right) = \sum_{i_{-1}, \ldots, i_{-n}} P\{x_{-1} = i_{-1}, \ldots, x_{-n} = i_{-n}\} \times$$
$$\times \sum_{i_0} \eta(P\{x_0 = i_0 \mid x_{-1} = i_{-1}, \ldots, x_{-n} = i_{-n}\}).$$

Отсюда вытекает, что энтропия сдвига удовлетворяет условию

$$h(T) \leqslant \ln r, \tag{7.1}$$

где r — число состояний.

Если T — сдвиг Бернулли $(p_1, \ldots, p_r)$, то

$$P\{x_0 = i_0 \mid x_{-1} = i_{-1}, \ldots, x_{-n} = i_{-n}\} = p_i$$

и все $H\left(\mathcal{A} \mid \bigvee_{k=1}^{n} T^k \mathcal{A}\right)$ совпадают с $-\sum_i p_i \ln p_i$.

Итак,

$$h(T) = h(\mathcal{A}, T) = -\sum_i p_i \ln p_i. \tag{7.2}$$

Пусть теперь T — сдвиг Маркова, соответствующий матрице перехода $\Pi = (p_{ij})$ и стационарным вероятностям $p = (p_i)$. На этот раз $P\{x_0 = i_0 \mid x_{-1} = i_{-1}, \ldots, x_{-n} = i_{-n}\} = p_{i_{-1} i_0}$, так что

$$H\left(\mathcal{A} \mid \bigvee_{k=1}^{n} T^k \mathcal{A}\right) = -\sum_{i_{-1}, \ldots, i_{-n}} p_{i_{-n}} p_{i_{-n} i_{-n+1}} \ldots$$
$$\ldots p_{i_{-2} i_{-1}} \sum_{i_0} p_{i_{-1} i_0} = -\sum_{ij} p_i p_{ij} \ln p_{ij}.$$

Поэтому

$$h(T) = h(\mathcal{A}, T) = -\sum_{ij} p_i p_{ij} \ln p_{ij}, \tag{7.3}$$

и это справедливо без каких бы то ни было предположений регулярности, относящихся к матрице Π.

Этот результат можно также установить, доказав что

$$H\left(\bigvee_{k=0}^{n-1} T^{-k} \mathcal{A}\right) = -(n-1) \sum_{ij} p_i p_{ij} \ln p_{ij} - \sum_i p_i \ln p_i \qquad (n \geqslant 1).$$

Разумеется, сдвиг Бернулли является также марковским сдвигом, и в этом случае (7.2) и (7.3) дают одинаковый результат. Таким образом, два сдвига Бернулли, или, более общо, два сдвига Маркова неизоморфны, если их энтропии различны. В частности, сдвиги Бернулли (1/2, 1/2) и

(1/3, 1/3, 1/3), имеющие энтропии $\ln 2$ и $\ln 3$ соответственно, неизоморфны.

Не представляет труда показать, что сдвиг примера 3.3 имеет энтропию $-\left(\frac{1}{2}\right)\sum_i p_i \ln p_i$.

В силу следствия теоремы 7.1 соотношения (7.2) и (7.3) справедливы для односторонних сдвигов Бернулли и Маркова.

Интересно, что если T обратимо и $\bigvee_{n=0}^{\infty} T^{-n}\mathcal{A} = \mathcal{F}$, то $h(T) = h(\mathcal{A}, T) = 0$. Интуитивно: в этом случае прошедшее определяет будущее или будущее определяет прошедшее; это зависит от того, какая выбрана ориентация, и, следовательно, соответствующая условная энтропия должна стремиться к 0.

Доказательство заключается в облечении этой идеи в математическую форму. В самом деле, из этого следует, что $\mathcal{A} \subset \mathcal{F} = T^{-1}\mathcal{F} = \bigvee_{i=1}^{\infty} T^{-i}\mathcal{A}$, так что по теореме 6.1 поле $\bigcup_{n=1}^{\infty} \bigvee_{i=1}^{n} T^{-i}\mathcal{A}$ содержит для каждого ε такое конечное подполе $\mathcal{B}$, что $H(\mathcal{A}\,|\,\mathcal{B}) < \varepsilon$. Но тогда $H\left(\mathcal{A}\,\middle|\,\bigvee_{i=1}^{n} T^{-i}\mathcal{A}\right) \leqslant H(\mathcal{A}\,|\,\mathcal{B}) < \varepsilon$ для большего n, так что в силу (B_1) $h(T) = h(\mathcal{A}, T) = 0$.

Опираясь на этот факт, можно показать, что энтропия иррационального поворота окружности (пример 1.5) равна 0. Пусть атомы поля $\mathcal{A}$ — верхняя полуокружность A и ее дополнение A^c. Если $T\omega = c\omega$, то полуокружность $T^{-n}A$ начинается в точке c^{-n}. Если c не является корнем из единицы, то множество $\{c^{-1}, c^{-2}, \ldots\}$ всюду плотно и, следовательно, *любая* полуокружность может быть аппроксимирована полуокружностями $T^{-n}A$ и, таким образом, лежит в $\bigvee_{n=0}^{\infty} T^{-n}\mathcal{A}$. Из этого вытекает, что $\bigvee_{n=0}^{\infty} T^{-n}\mathcal{A}$ содержит любую дугу и, следовательно, совпадает с $\mathcal{F}$.

Из соотношений (B_4) и (B_2) вытекает, что $h(T^k) = kh(T)$ для $k \geqslant 1$. Если T обратимо, то $h(T) = h(T^{-1})$, так как

$$H\left(\bigvee_{k=0}^{n-1} T^{-k}\mathcal{A}\right) = H\left(\bigvee_{k=0}^{n-1} (T^{-1})^{-k}\mathcal{A}\right).$$

Таким образом, мы можем вычислить энтропию любой степени преобразования T.

Теорема 7.2. *Если $k \geqslant 1$, то $h(T^k) = kh(T)$. Если T обратимо, то $h(T^k) = |k| \cdot h(T)$ для любого целого k.*

Если T — вращение окружности $T\omega = c\omega$, где c — корень из единицы, то $h(T) = 0$. Действительно, если c есть корень k-й степени из единицы, то $T^k = I$, так что $h(T) = k \cdot h(I) = 0$.

Некоторые обобщения[1])

Пусть $\mathscr{G}_1$ и $\mathscr{G}_2$ — σ-подполя поля $\mathscr{F}$; если каждое множество из $\mathscr{G}_1$ отличается от некоторого множества из $\mathscr{G}_2$ на множество меры 0 и обратно, то будем писать $\mathscr{G}_1 \overset{*}{=} \mathscr{G}_2$. Следующая теорема содержит в числе других результатов теорему 7.1.

Теорема 7.3. *Пусть $\{\mathscr{G}_n\}$ — неубывающая последовательность (конечно аддитивных) полей. Если*

$$\bigvee_{n=1}^{\infty} \bigvee_{i=0}^{\infty} T^{-i}\mathscr{G}_n \overset{*}{=} \mathscr{F}$$

или если T обратимо и

$$\bigvee_{n=1}^{\infty} \bigvee_{i=-\infty}^{\infty} T^{i}\mathscr{G}_n \overset{*}{=} \mathscr{F},$$

то

$$h(T) = \lim_{n\to\infty} \sup_{\mathscr{A} \subset \mathscr{G}_n} h(\mathscr{A}, T).$$

Доказательство. Дадим доказательство для случая $\bigvee_{n=1}^{\infty} \bigvee_{i=0}^{\infty} T^{-i}\mathscr{G}_n \overset{*}{=} \mathscr{F}$. Если $\mathscr{H}_n$ — поле, порожденное $\bigcup_{i=0}^{n} T^{-i}\mathscr{G}_n$, и $\mathscr{F}_0$ — поле $\bigcup_{n=1}^{\infty} \mathscr{H}_n$, то каждое множество поля $\mathscr{F}$ отличается от некоторого множества σ-поля, порожденного полем $\mathscr{F}_0$, на множество меры 0, откуда в силу теоремы 6.1 и соотношения (B_5) следует (так же как в центральном пункте доказательства теоремы 7.1), что

$$h(T) = \sup_{\mathscr{B} \subset \mathscr{F}_0} h(\mathscr{B}, T). \tag{7.4}$$

---

[1]) Изложенные в этом разделе результаты в дальнейшем использованы не будут.

Если $\mathscr{B} \subset \mathscr{F}_0$, то $\mathscr{B}$ содержится в $\mathscr{H}_n$ при некотором n и, следовательно [1]), имеет атомы $B_1, \ldots, B_k$ вида

$$B_u = \bigcup_{v=1}^{l} \bigcap_{i=0}^{n} T^{-i} G_{iuv}, \quad u = 1, \ldots, k,$$

где $G_{iuv} \in \mathscr{G}_n$. Если $\mathscr{A}$ — поле, порожденное множествами G_{iuv}, то оно конечно, $\mathscr{A} \subset \mathscr{G}_n$ и $\mathscr{B} \subset \bigvee_{i=0}^{\infty} T^{-i}\mathscr{A}$. Поэтому в силу ($B_2$) и ($B_3$) имеем

$$h(\mathscr{B}, T) \leqslant h\left(\bigvee_{i=0}^{n} T^{-i}\mathscr{A}, T\right) = h(\mathscr{A}, T) \leqslant \sup_{\mathscr{A} \subset \mathscr{G}_n} h(\mathscr{A}, T).$$

Так как правый член этого неравенства не убывает по n, то теорема следует из (7.4). Случай обратимого T разбирается аналогично.

Если $\mathscr{G}_n$ есть σ-поле и $T^{-1}\mathscr{G}_n \subset \mathscr{G}_n$, то T может рассматриваться как сохраняющее меру преобразование T_n на пространстве $(\Omega, \mathscr{G}_n, P)$. Если это так, то из теоремы 7.3 следует, что если последовательность $\{\mathscr{G}_n\}$ не убывает и $\bigvee_{n=1}^{\infty} \mathscr{G}_n \stackrel{*}{=} \mathscr{F}$, то $h(T) = \lim_n h(T_n)$.

Взяв в теореме 7.3 в качестве $\mathscr{G}_n$ поле $\mathscr{A}$, получаем как следствие теорему 7.1.

Следствие 1. *Если $\{\mathscr{A}_n\}$ — неубывающая последовательность конечных полей, причем $\bigvee_{n=1}^{\infty} \mathscr{A}_n \stackrel{*}{=} \mathscr{F}$, то*

$$h(T) = \lim_{n \to \infty} h(\mathscr{A}_n, T).$$

Следствие 2. *Если $\mathscr{B} \subset \bigvee_{i=0}^{\infty} T^{-i}\mathscr{A}$ или если T обратимо и $\mathscr{B} \subset \bigvee_{i=-\infty}^{\infty} T^{i}\mathscr{A}$, то $h(\mathscr{B}, T) \leqslant h(\mathscr{A}, T)$.*

Для доказательства этого следствия рассмотрим T как преобразование на $(\Omega, \mathscr{F}_0, P_0)$, где $\mathscr{F}_0$ — поле $\bigvee_{i=0}^{\infty} T^{-i}\mathscr{A}$ или $\bigvee_{i=-\infty}^{\infty} T^{i}\mathscr{A}$ и P_0 — сужение P на $\mathscr{F}_0$.

[1]) Если $\mathscr{M}$ и $\mathscr{N}$ — поля, то поле, порожденное полем $\mathscr{M} \cdot \bigcup \mathscr{N}$, состоит из конечных объединений множеств вида $M \cap N$, где $M \in \mathscr{M}$ и $N \in \mathscr{N}$.

Следствие 3. *Если* $\mathscr{G}$ — *некоторое поле и либо* $\bigvee_{i=0}^{\infty} T^{-i}\mathscr{G} \stackrel{*}{=} \mathscr{F}$, *либо* T *обратимо и* $\bigvee_{i=-\infty}^{\infty} T^{i}\mathscr{G} \stackrel{*}{=} \mathscr{F}$, *то*

$$h(T) = \sup_{\mathscr{A} \subset \mathscr{G}} h(\mathscr{A}, T).$$

Следствие 4. *Если* $\mathscr{F}_0$ — *поле, порождающее* $\mathscr{F}$, *то*

$$h(T) = \sup_{\mathscr{A} \subset \mathscr{F}_0} h(\mathscr{A}, T).$$

Это последнее следствие дает возможность вычислять энтропию прямых произведений. Пусть T_i $(i = 1, 2)$ — сохраняющее меру преобразование на пространстве $(\Omega_i, \mathscr{F}_i, P_i)$; тогда прямое произведение $T_1 \times T_2$ — сохраняющее меру преобразование на произведении этих двух пространств:

$$(T_1 \times T_2)(\omega_1, \omega_2) = (T_1\omega_1, T_2\omega_2).$$

Теорема 7.4. *Прямое произведение преобразований* T_1 *и* T_2 *обладает свойством*

$$h(T_1 \times T_2) = h(T_1) + h(T_2).$$

Доказательство. Если $\mathscr{A}_i$ — конечное подполе поля $\mathscr{F}_i$ $(i = 1, 2)$, то, обозначив $2_i = \{0, \Omega_i\}$, имеем

$$\bigvee_{i=0}^{n-1} (T_1 \times T_2)^{-i}(\mathscr{A}_1 \times \mathscr{A}_2) = \bigvee_{i=0}^{n-1} (T_1^{-i}\mathscr{A}_1 \times 2_2) \vee \bigvee_{i=0}^{n-1} (2_1 \times T_2^{-i}\mathscr{A}_2).$$

(Если $\mathscr{A}_1$ и $\mathscr{A}_2$ — это σ-поля в пространствах Ω_1 и Ω_2, то $\mathscr{A}_1 \times \mathscr{A}_2$ есть σ-поле в $\Omega_1 \times \Omega_2$, порожденное $A_1 \times A_2$, где A_1 и A_2 лежат в $\mathscr{A}_1$ и $\mathscr{A}_2$ соответственно. Если $\mathscr{A}_1$ и $\mathscr{A}_2$ — конечные поля, то поле $\mathscr{A}_1 \times \mathscr{A}_2$, также конечно и его атомами являются множества $A_1 \times A_2$, где A_1 и A_2 — атомы полей $\mathscr{A}_1$ и $\mathscr{A}_2$ соответственно. Два поля в правой части этого равенства независимы в том смысле, что если M и N принадлежат соответственно этим полям, то $P(M \cap N) = P(M)P(N)$, где $P = P_1 \times P_2$; отсюда следует, что

$$H\left(\bigvee_{i=0}^{n-1} (T_1 \times T_2)^{-i}(\mathscr{A}_1 \times \mathscr{A}_2)\right) =$$

$$= H\left(\bigvee_{i=0}^{n-1} (T_1^{-i}\mathscr{A}_1 \times 2_2)\right) + H\left(\bigvee_{i=0}^{n-1} (2_1 \times T_2^{-i}\mathscr{A}_2)\right) =$$

$$= H\left(\bigvee_{i=0}^{n-1} T_1^{-i}\mathscr{A}_1\right) + H\left(\bigvee_{i=0}^{n-1} T_2^{-i}\mathscr{A}_2\right).$$

Деля на n и переходя к пределу, получаем

$$h(\mathcal{A}_1 \times \mathcal{A}_2,\ T_1 \times T_2) = h(\mathcal{A}_1,\ T_1) + h(\mathcal{A}_2,\ T_2).$$

Теперь теорема вытекает из следствия 4, так как прямоугольники с основаниями в $\mathcal{F}_1$ и $\mathcal{F}_2$ порождают поле $\mathcal{F}_1 \times \mathcal{F}_2$.

В качестве последнего вычисления найдем энтропию примера 1.7 с мерой P, определенной как произведение мер по формуле

$$P\{\omega:\ x_l(\omega) \in E_l,\ n \leqslant l \leqslant n+k\} = \prod_{l=n}^{n+k} \mu(E_l), \tag{7.5}$$

где μ есть мера на борелевских подмножествах прямой. Из следствия 3 вытекает

$$h(T) = \sup \sum_{i=1}^{k} \eta(\mu(E_i)),$$

где верхняя грань берется по всем разбиениям $E_1, \ldots, E_k$ прямой на борелевские множества. Если μ не состоит исключительно из точечных масс, то $h(T) = \infty$. Если μ состоит из точечных масс, скажем, $p_1, p_2, \ldots$, то

$$h(T) = -\sum_{j=1}^{\infty} p_j \ln p_j.$$

Если этот ряд расходится, то $h(T)$ бесконечна.

Замечание. Материал этого и предыдущего параграфов почерпнут из работ Халмоша [4] и Брауна [1], а также из статей Колмогорова и Синая. Дальнейшую информацию по этому поводу можно получить из превосходной обзорной статьи Рохлина [2]. Статьи более позднего времени включены в библиографию, помещенную в конце книги.

8. ПРОБЛЕМА ПОЛНОТЫ [1])

Некоторые нерешенные задачи

В § 5 мы отметили, что если T_k — циклическая перестановка k точек равной массы, то $h(T_k) = 0$, хотя различные T_k неизоморфны. Кроме того, энтропия любого вращения окружности также равна 0, хотя оно неизоморфно никакой

[1]) Материал этого параграфа, который состоит преимущественно из вопросов, встречается в последующем изложении лишь эпизодически.

перестановке T_k. Энтропия, таким образом, не является полным инвариантом.

Если T — сдвиг Маркова, то, как мы видели,

$$h(T) = -\sum_{ij} p_i p_{ij} \ln p_{ij}.$$

Рассмотрим три матрицы перехода

$$\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

и соответствующие стационарные распределения вероятностей

$$\left(\frac{1}{2}, \frac{1}{2}\right), \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right), \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right);$$

пусть T_1, T_2 и T_3 — соответствующие сдвиги Маркова. Тогда T_1 — перемешивающий сдвиг, T_2 — эргодический, но не перемешивающий, а T_3 — даже не эргодический; в частности, никакие два из этих сдвигов неизоморфны. Так как энтропия всех сдвигов равна $\ln 2$, то она не является полным инвариантом и для сдвигов Маркова.

Общий вопрос об условиях, при которых энтропия является полным инвариантом, порождает несколько специальных неразрешенных вопросов.

Вопрос 1. *Является ли энтропия полным инвариантом для сдвигов Бернулли?* Пусть T — сдвиг Бернулли $(p_1, \ldots, p_r)$, а $\tilde{T}$ — сдвиг Бернулли $(\tilde{p}_1, \ldots, \tilde{p}_{\tilde{r}})$. Если T и $\tilde{T}$ изоморфны, то

$$-\sum_i p_i \ln p_i = -\sum_i \tilde{p}_i \ln \tilde{p}_i. \tag{8.1}$$

Следует ли из (8.1), что T и $\tilde{T}$ изоморфны? Ответ неизвестен. Если $(p_1, \ldots, p_r)$ — некоторая перестановка $(\tilde{p}_1, \ldots, \tilde{p}_{\tilde{r}})$ $(r = \tilde{r})$, то ответ положителен — это тривиально. Однако возьмем наборы вероятностей (1/4, 1/4, 1/4, 1/4) и (1/2, 1/8, 1/8, 1/8, 1/8). Тогда (8.1) выполняется, но a priori у нас не больше оснований ожидать изоморфность T и $\tilde{T}$, чем изоморфность

сдвигов Бернулли (1/2, 1/2) и (1/3, 1/3, 1/3). Мешалкин доказал поразительный факт: T и $\tilde{T}$ изоморфны. Это позволяет думать, что ответ на вопрос 1 положителен.

В этом направлении Синай доказал следующую замечательную теорему. Сказать, что два сохраняющих меру преобразования T и $\tilde{T}$ изоморфны, значит сказать (пренебрегая множествами меры 0), что существует взаимно однозначное отображение φ пространства Ω на $\tilde{\Omega}$, которое переводит меру P в $\tilde{P}$ и удовлетворяет условию $\tilde{T}\varphi = \varphi T$. Пусть мы требуем, чтобы областью значений отображения φ было все $\tilde{\Omega}$, чтобы φ переводило P в $\tilde{P}$ и удовлетворяло условию $\tilde{T}\varphi = \varphi T$, но не требуем, чтобы оно было взаимно однозначным. Если такое отображение существует, говорят, что $\tilde{T}$ есть *факторпреобразование* преобразования T. Легко показать в этом случае, что $h(\tilde{T}) \leqslant h(T)$. Если T и $\tilde{T}$ являются факторпреобразованиями друг друга, то $h(T) = h(\tilde{T})$; тогда говорят, что они *слабо изоморфны*. (Можно показать, что слабый изоморфизм сохраняет не только энтропию, но и свойства эргодичности и перемешивания.) Синай показал, что два сдвига Бернулли с одинаковой энтропией слабо изоморфны и отображения $\varphi\colon \Omega \to \tilde{\Omega}$ и $\psi\colon \tilde{\Omega} \to \Omega$, входящие в определение слабого изоморфизма, могут быть выбраны так, чтобы они зависели только от прошлого. Это означает, что n-я координата $(\varphi\omega)_n$ $[(\psi\tilde{\omega})_n]$ точки $\varphi\omega$ $[\psi\tilde{\omega}]$ является функцией координат $\ldots, \omega_{n-1}, \omega_n$ $[\ldots, \tilde{\omega}_{n-1}, \tilde{\omega}_n]$ точки ω $[\tilde{\omega}]$.

Хотя обратное тривиально, но неизвестно, когда слабый изоморфизм влечет обычный изоморфизм. Для того чтобы показать заключенную здесь трудность, рассмотрим такой пример. Пусть $T = \tilde{T}$ — двусторонний сдвиг Бернулли (1/2, 1/2) с пространством состояний $\{0, 1\}$. Тогда T и $\tilde{T}$ слабо изоморфны: соответствующие отображения φ и ψ определяются соотношениями

$$(\varphi\omega)_n = \omega_{n-1} + \omega_n \pmod 2,$$

$$(\psi\tilde{\omega})_n = \tilde{\omega}_{n-1} + \tilde{\omega}_n \pmod 2.$$

Разумеется, будучи тождественными, T и $\tilde{T}$ изоморфны. Но каждое из отображений φ и ψ является однозначным отображением, при котором точке соответствуют два прообраза, и весьма трудно понять, как их можно использовать для того, чтобы восстановить тождественное или любое другое обратимое отображение, которое делает T и $\tilde{T}$ изоморфными.

Вопрос 2. *Является ли энтропия полным инвариантом для перемешивающих сдвигов Маркова?* Ответ на этот вопрос также неизвестен.

Сдвиги Колмогорова

Прежде чем поставить наш третий вопрос, введем новое понятие. Пусть $\mathcal{A}$ — поле событий, наблюдавшихся в момент времени 0, для общего сдвига с пространством состояний ρ (пример 1.2). Если $\mathcal{F}_n = \bigvee_{k=n}^{\infty} T^{-k}\mathcal{A}$, то $\mathcal{F}_n$ есть σ-поле, порожденное множествами $\{x_k = i\}$, где $i \in \rho$ и $k \geqslant n$. Сдвиг T называется сдвигом Колмогорова, если все множества σ-поля $\mathcal{F}_\infty = \bigcap_{n=1}^{\infty} \mathcal{F}_n$ тривиальны, т. е. имеют меру либо 0, либо 1. Множества из $\mathcal{F}_\infty$ зависят от „сколь угодно далекого будущего“ (см. конец § 4, где подобный класс множеств определен для преобразования, связанного с непрерывными дробями).

Существует общее понятие преобразования Колмогорова [1]), определяемое инвариантным образом и приводящее к упомянутому выше понятию в случае сдвига.

В § 11 мы увидим, что любой сдвиг Колмогорова — перемешивающий. Можно построить пример, показывающий, что не все сдвиги, обладающие свойством перемешивания, являются сдвигами Колмогорова.

Из закона нуля — единицы следует, что любой сдвиг Бернулли является сдвигом Колмогорова. Следующее рассуждение показывает, что сдвиг, соответствующий неприводимой непериодической цепи Маркова, является сдвигом Колмогорова. Действительно, в этом случае $\lim\limits_n p_{ij}^{(n)} = p_j$, так что если $\varepsilon_n = \max\limits_{i,\, j \in \rho} \left| p_{ij}^{(n)} - p_j \right|$, то $\lim\limits_n \varepsilon_n = 0$. Из теории цепей Маркова следует, что если множество A принадлежит σ-полю, порожденному множествами $\{x_k = i\}$, где $k \leqslant l$, а множество B принадлежит σ-полю, порожденному множествами $\{x_k = i\}$, где $k \geqslant l + n$ (т. е. полю $\mathcal{F}_{l+n}$), то

$$|P(A \cap B) - P(A)P(B)| \leqslant \varepsilon_n. \tag{8.2}$$

---

[1]) В отечественной литературе принят термин *K-автоморфизм*. Определение *K*-автоморфизма см. в сноске 2 на стр. 111. — *Прим. ред.*

Если B принадлежит полю $\mathscr{F}_\infty$, то оно принадлежит и полю $\mathscr{F}_{l+n}$ при всех n, так что для любого цилиндра A

$$P(A \cap B) = P(A)\,P(B). \tag{8.3}$$

Но совокупность множеств A, для которых справедливо (8.3), образует σ-поле и, следовательно, совпадает с $\mathscr{F}$. Таким образом, (8.3) имеет место для $A = B$, так что $P(B)$ равняется либо 0, либо 1.

Вопрос 3. *Является ли энтропия полным инвариантом для сдвигов Колмогорова*? Ответ неизвестен.

В качестве последнего вопроса рассмотрим следующую проблему изоморфизма, которую одна энтропия разрешить не может. Пусть T и $\tilde{T}$ — два варианта преобразования примера 1.7, соответствующие произведениям двух различных мер μ и $\tilde{\mu}$ на прямой (см. (7.5)). Если ни та, ни другая меры не состоят исключительно из точечных масс, то $h(T) = h(\tilde{T}) = \infty$. При каких условиях в этом случае T и $\tilde{T}$ изоморфны? Если меры μ и $\tilde{\mu}$ совсем не имеют точечных масс, то хорошо известно, что существует отображение прямой на саму себя, переводящее μ в $\tilde{\mu}$. Это отображение можно расширить так, чтобы получить изоморфизм между T и $\tilde{T}$. А что можно сказать о смешанном случае?

Замечание. По поводу изоморфизма сдвигов Бернулли (1/4, 1/4, 1/4, 1/4) и (1/2, 1/8, 1/8, 1/8) см. Мешалкин [1]. О понятии слабого изоморфизма см. Синай [8] [1]). Колмогоров [2] ввел преобразование Колмогорова, назвав его квазирегулярным [2]).

---

[1]) См. также Синай [9]. — *Прим. ред.*

[2]) Общее определение таково: обратимое сохраняющее меру преобразование T пространства $(\Omega, \mathscr{F}, P)$ называется *автоморфизмом Колмогорова* (или *К-автоморфизмом*), если существует σ-поле $\mathscr{F}_0 \subset \mathscr{F}$ со свойствами: 1) $T\mathscr{F}_0 \supset \mathscr{F}_0$, 2) $\bigvee\limits_n T^n\mathscr{F}_0 = \mathscr{F}$, 3) $\bigwedge\limits_n T^n\mathscr{F}_0$ тривиально. В случае сдвига в пространстве последовательностей в качестве $\mathscr{F}_0$ можно взять σ-поле, порожденное величинами x_i при $i \leqslant 0$ или при $i \geqslant 0$. — *Прим. ред.*

ГЛАВА 3

Условные вероятности и математические ожидания[1]

9. УСЛОВНЫЕ ВЕРОЯТНОСТИ

Конечный случай

Понятие условных вероятностей относительно σ-поля лежит в основе многих построений современной теории вероятностей. Рассмотрим сначала понятие условной вероятности множества M относительно другого множества A. Эта вероятность определяется равенством $P(M \mid A) = P(M \cap A)/P(A)$ (если только $P(A)$ не равна нулю, в последнем случае она вообще не определена).

Если Тихе выбирает точку ω из пространства Ω в соответствии с вероятностной мерой P (на σ-поле $\mathscr{F}$), то $P(M)$ есть вероятность того, что ω взята из M. Если эта выбранная точка принадлежит A и если Тихе сообщает об этом (и только об этом) экспериментатору, то для него $P(M \mid A)$ есть вероятность того, что ω принадлежит также и M. Будем исходить из этой эвристической формулировки.

С другой стороны, если точка, выбранная Тихе, оказалась принадлежащей A^c и экспериментатору сообщено об этом, то для него новая вероятность события $\omega \in M$ равна $P(M \mid A^c)$. Удобно связать эти две условные вероятности ступенчатой функцией

$$f(\omega) = \begin{cases} P(M \mid A) = \dfrac{P(M \cap A)}{P(A)}, & \text{если } \omega \in A, \\ P(M \mid A^c) = \dfrac{P(M \cap A^c)}{P(A^c)}, & \text{если } \omega \in A^c. \end{cases}$$

Если Тихе сообщает экспериментатору, какому множеству — A или A^c — принадлежит выбранная ею точка, то вероятность события $\omega \in M$ для экспериментатора равна $f(\omega)$. Хотя экспериментатор не знает аргумента ω функции f, он бла-

---

[1] Читатель, знакомый с понятиями условных вероятностей и математических ожиданий относительно σ-поля, может опустить эту главу, при необходимости находя в ней нужные ссылки. Читатель, незнакомый с этими идеями, также может опустить ее, если он согласен принять несколько дальнейших теорем без доказательства или если он будет предполагать, что встречающиеся в этих теоремах сдвиги являются марковскими.

годаря Тихе достаточно осведомлен для вычисления $f(\omega)$. Заметим, что значение $f(\omega)$ определяет, будет ли $\omega \in A$ или $\omega \in A^c$, если только не выполняется равенство $P(M \mid A) = P(M \mid A^c)$ (означающее независимость M и A), вследствие которого условная вероятность множества M совпадает с безусловной. Поэтому вместо того, чтобы сообщать, принадлежит ω множеству A или A^c, Тихе могла бы сообщать значение $f(\omega)$.

Множества A и A^c образуют $\mathcal{F}$-разбиение пространства Ω. Только что изложенное приводит нас к произвольному $\mathcal{F}$-разбиению, или, что то же, к произвольному конечному подполю $\mathcal{A}$ поля $\mathcal{F}$. Если $A_1, \ldots, A_r$ являются атомами подполя $\mathcal{A}$, то для любого M из $\mathcal{F}$ рассмотрим ступенчатую функцию

$$f(\omega) = P(M \mid A_i) = \frac{P(M \cap A_i)}{P(A_i)}, \text{ если } \omega \in A_i, \quad i = 1, \ldots, r.$$

Тихе, выбрав точку ω из Ω в соответствии с мерой P, говорит экспериментатору, какой из атомов подполя $\mathcal{A}$ ее содержит. Новая вероятность множества M для экспериментатора имеет значение $f(\omega)$. Как и раньше, экспериментатор после сообщения Тихе имеет достаточные данные для вычисления $f(\omega)$, даже если само ω ему неизвестно. И снова знание $f(\omega)$ равносильно знанию того, какое из A_i содержит ω.

Будем обозначать через $P\{M \| \mathcal{A}\}$ функцию (случайную величину) f и называть ее *условной вероятностью множества M при заданном конечном поле $\mathcal{A}$*. Аргумент ω обычно опускается, но в случае необходимости будем писать $P\{M \| \mathcal{A}\}_\omega$.

Таким образом, $P\{M \| \mathcal{A}\}$ — ступенчатая функция, значение которой на A_i равно обычной условной вероятности $P(M \mid A_i)$. Это определение требует дополнения, так как $P(M \mid A_i)$ не определена при $P(A_i) = 0$. Мы можем оставить $P\{M \| \mathcal{A}\}$ неопределенной на таких атомах поля $\mathcal{A}$, а можем приписывать ей какое-то фиксированное значение, скажем, 0. Выберем третий путь: если $P(A_i) = 0$, то значение $P\{M \| \mathcal{A}\}$ на A_i есть произвольное действительное число [1]). Итак, если $\mathcal{A}$ содержит атомы меры 0, то $P\{M \| \mathcal{A}\}$ является для них целым семейством функций на Ω. Чтобы подчеркнуть это, мы будем называть $P\{M \| \mathcal{A}\}$ *вариантом* условной

[1]) Мы требуем, чтобы это значение было одним и тем же для всех точек A_i, так что $P\{M \| \mathcal{A}\}$ есть ступенчатая функция, для которой атомы поля $\mathcal{A}$ являются множествами ее постоянства.

вероятности множества M при заданном $\mathcal{A}$. Заметим, что любые два варианта равны почти всюду.

Любой вариант случайной величины $P\{M\|\mathcal{A}\}$ обладает двумя основными свойствами: 1) эта функция интегрируема и измерима относительно $\mathcal{A}$[1]) и 2) равенство

$$\int_A P\{M\|A\}\,dP = P(M \cap A) \tag{9.1}$$

справедливо для любого множества A из конечного поля $\mathcal{A}$. Легко показать, что любая случайная величина, обладающая этими двумя свойствами, является вариантом функции $P\{M\|\mathcal{A}\}$.

Общий случай

Так как $\mathcal{A}$ состоит из конечных объединений непересекающихся атомов $A_1, \ldots, A_r$, то знание того, какой из атомов поля $\mathcal{A}$ содержит ω, равносильно знанию того, какое из 2^r множеств поля $\mathcal{A}$ содержит ω, а какое нет. Этот второй подход переносится на произвольное σ-подполе поля $\mathcal{F}$.

Если $\mathcal{G}$ — такое σ-подполе, то можно представить себе, что Тихе выбирает точку ω и затем открывает экспериментатору, какие из множеств G поля $\mathcal{G}$ содержат ω, а какие нет[2]). Определим функцию $P\{M\|\mathcal{G}\}$, значение $P\{M\|\mathcal{G}\}_\omega$ которой в точке ω является тем, что, как мы интуитивно чувствуем, должно быть для экспериментатора новой вероятностью события $\omega \in M$ после получения сообщения Тихе.

Для любого M из $\mathcal{F}$ определим $P\{M\|\mathcal{G}\}$ как интегрируемую случайную величину, которая a) измерима относительно σ-подполя $\mathcal{G}$ и b) удовлетворяет функциональному уравнению

$$\int_G P\{M\|\mathcal{G}\}\,dP = P(M \cap G), \qquad G \in \mathcal{G}. \tag{9.2}$$

Если $\mathcal{G}$ конечно, то это определение сводится к предыдущему (см. 9.1).

Условие (a) соответствует требованию, чтобы $P\{M\|\mathcal{G}\}$ можно было вычислить на основании только одного сооб-

---

[1]) Для этого мы требуем, чтобы $P\{M\|\mathcal{A}\}$ была постоянна даже на атомах меры 0.

[2]) Тихе предъявляет не ω, а вектор с компонентами $I_G(\omega)$, отмеченными индексами G из $\mathcal{G}$.

щения Тихе. Условие (b) допускает игровую интерпретацию. Предположим, что Тихе, передав экспериментатору свое сообщение, предоставляет ему возможность поставить на наступление события M (если M не лежит в $\mathscr{G}$, он не знает, наступило это событие или нет). Она требует выплаты начальной ставки и берется выплачивать одну драхму, если событие M наступило, и ничего — в противном случае. Если условная вероятность события M для экспериментатора после сообщения Тихе равна $P\{M \| \mathscr{G}\}$, то справедливый размер начальной ставки равен $P\{M \| \mathscr{G}\}$ драхм. Если экспериментатор решит принять пари и внесет эту ставку, то он выиграет $1 - P\{M \| \mathscr{G}\}$ в случае наступления M и $-P\{M \| \mathscr{G}\}$ в противном случае, так что его выигрыш равен

$$(1 - P\{M \| \mathscr{G}\})I_M + (-P\{M \| \mathscr{G}\})I_{M^c} = I_M - P\{M \| \mathscr{G}\}. \quad (9.3)$$

Если он отклоняет пари, то его выигрыш автоматически равен 0. Предположим, что он придерживается стратегии принимать пари, если наступает событие G и отклонять его в противном случае. (G здесь есть некоторое множество из $\mathscr{G}$.) Его ожидаемый выигрыш при этой стратегии равен интегралу от (9.3) по множеству G:

$$\int_G (I_M - P\{M \| \mathscr{G}\})\, dP. \quad (9.4)$$

Но (9.2) является в точности требованием того, что (9.4) равно нулю для каждого G из $\mathscr{G}$. Если $P\{M \| \mathscr{G}\}$ удовлетворяет нашим эвристическим требованиям, то справедливая начальная ставка не приведет ни к выгодной, ни к невыгодной для экспериментатора стратегии.

Нужно, разумеется, доказать существование случайных величин $P\{M \| \mathscr{G}\}$, удовлетворяющих условиям (a) и (b), что мы сделаем, применяя теорему Радона — Никодима. Определим вполне аддитивную функцию φ множества на σ-поле $\mathscr{G}$ равенством

$$\varphi(G) = P(M \cap G), \qquad G \in \mathscr{G}.$$

Тогда φ есть конечная мера на $\mathscr{G}$, которая, очевидно, абсолютно непрерывна относительно меры P, рассматриваемой на $\mathscr{G}$ (для $G \in \mathscr{G}$ из $P(G) = 0$ вытекает $\varphi(G) = 0$). Следовательно, в силу теоремы Радона — Никодима существует интегрируемая функция f, измеримая относительно $\mathscr{G}$ и

удовлетворяющая условию [1])

$$\int_G f\,dP = \varphi(G), \qquad G \in \mathscr{G}. \tag{9.5}$$

Возьмем в качестве f функцию $P\{M \| \mathscr{G}\}$ и назовем ее условной вероятностью события M при заданном $\mathscr{G}$; тогда (9.5) принимает вид (9.2). Таким образом, функция, определяющая условную вероятность, существует.

Может существовать и более одной такой случайной величины $P\{M \| \mathscr{G}\}$. Мы видели, что это может быть так даже для конечного $\mathscr{G}$. Тогда, как и раньше, говорят о некотором *варианте* условной вероятности. Любые два таких варианта f и g интегрируемы и измеримы относительно $\mathscr{G}$ и удовлетворяют соотношению $\int_G f\,dP = \int_G g\,dP$ для всех G из $\mathscr{G}$; отсюда следует, что $f = g$ п. в. С точностью до этого обстоятельства условные вероятности определены единственным образом [2]).

Пример 9.1. Предположим, что $M \in \mathscr{G}$, что всегда имеет место, если, например, $\mathscr{G}$ совпадает со всем σ-полем $\mathscr{F}$. Тогда $P\{M \| \mathscr{G}\} = I_M$ п. в. Интуитивно, если $M \in \mathscr{G}$, то знание того, какой элемент из $\mathscr{G}$ содержит ω, означает, в частности, знание того, наступило или не наступило событие M.

Пример 9.2. Если $\mathscr{G} = 2 = \{0, \Omega\}$, то $P\{M \| \mathscr{G}\}_\omega = P(M)$ для любой точки ω. Из сообщения Тихе экспериментатор не узнает ничего такого, чего бы он не знал раньше.

Пример 9.3. Пусть Ω — плоскость R^2 и $\mathscr{F}$ — класс борелевских множеств на этой плоскости. Любая точка ω из Ω есть пара $(x(\omega), y(\omega))$, x и y — координатные переменные. Пусть σ-подполе $\mathscr{G}$ состоит из вертикальных полос, т. е.

---

[1]) Мы применяем здесь теорему Радона — Никодима не в исходном пространстве с мерой $(\Omega, \mathscr{F}, P)$, а в пространстве $(\Omega, \mathscr{G}, P)$. Это обеспечивает измеримость f относительно $\mathscr{G}$. Хотя $\int_G f\,dP$ есть тогда интеграл от f в смысле $(\Omega, \mathscr{G}, P)$, нетрудно показать, что он совпадает с интегралом в смысле $(\Omega, \mathscr{F}, P)$.

[2]) Наше определение слегка отличается от данного Дубом [1]. Он допускает в качестве варианта любую функцию $P\{M \| \mathscr{G}\}$, измеримую относительно $\mathscr{F}$, интегрируемую, удовлетворяющую функциональному уравнению (9.2) и равную п. в. некоторой функции, измеримой относительно $\mathscr{G}$. Если $\mathscr{G}$ конечно, это равнозначно отказу от требования, чтобы $P\{M \| \mathscr{G}\}$ была постоянна на атомах, имеющих меру 0.

элементами поля $\mathscr{G}$ являются прямые произведения $E\times R^1 = \{\omega\colon x(\omega)\in E\}$, где E — борелевское множество на прямой. Если экспериментатор знает для каждой вертикальной полосы $E\times R^1$, содержит она ω или нет, то он знает это, в частности, для каждой полосы $\{\xi\}\times R^1$, где ξ — действительное число, и обратно. Таким образом, сообщение Тихе эффективно дает значение $x(\omega)$. Предположим теперь, что P — вероятностная мера на $\mathscr{F}$, имеющая плотность $p(\xi, \eta)$ относительно плоской меры Лебега:

$$P(M) = \int_M\int p(\xi, \eta)\, d\xi\, d\eta.$$

Покажем, что если $M = \{\omega\colon y(\omega)\in F\}$, где F — линейное борелевское множество, то вариантом функции $P(M\,\|\,\mathscr{G})$ является $\psi(x(\omega))$, где

$$\psi(\xi) = \frac{\int_F p(\xi, \eta)\, d\eta}{\int_{R^1} p(\xi, \eta)\, d\eta}. \tag{9.6}$$

Так как $\psi(x(\omega))$ — измеримая функция от $x(\omega)$, она измерима относительно $\mathscr{G}$. Элемент $\mathscr{G}$ имеет вид $\{\omega\colon x(\omega)\in E\}$, поэтому нужно только показать, что

$$\int_{\{x\in E\}} \psi(x(\omega))\, P(d\omega) = P\{x\in E,\ y\in F\}. \tag{9.7}$$

Однако так как случайная величина x имеет распределение на прямой с плотностью $p(\xi) = \int_{R^1} p(\xi, \eta)\, d\eta$, то левая часть равенства (9.7) равна

$$\int_E p(\xi)\psi(\xi)\, d\xi = \int_E \left\{\int_F p(\xi, \eta)\, d\eta\right\} d\xi =$$

$$= \int_{E\times F}\int p(\xi, \eta)\, d\xi\, d\eta = P\{x\in E,\ y\in F\},$$

где предпоследнее равенство следует из теоремы Фубини. Таким образом, $\psi(x(\omega))$ есть вариант функции $P\{M\,\|\,\mathscr{G}\}$.

Правая часть равенства (9.6) является классической формулой для условной вероятности события $\{y\in F\}$ при условии $x = \xi$. Наше обсуждение может рассматриваться как оправдание этой классической формулы или как

пример, оправдывающий наше математическое определение условной вероятности.

Если x — случайная величина на вероятностном пространстве $(\Omega, \mathscr{F}, P)$, то σ-поле $\mathscr{G}$, порожденное ею (наименьшее σ-поле, относительно которого величина x измерима), состоит из множеств вида $\{\omega: x(\omega) \in E\}$, где E пробегает различные линейные борелевские множества. Так как знание того, какие множества этого вида содержат ω, а какие нет, равносильно знанию значения $x(\omega)$, назовем $P\{M \| \mathscr{G}\}$ условной вероятностью события M при заданном x и будем обозначать ее $P\{M \| x\}$. Пример 9.3 представляет собой частный случай. Таким же образом определим условную вероятность $P\{M \| x_1, x_2, \ldots\}$ события M при заданной конечной или бесконечной последовательности величин $x_1, x_2, \ldots$ как $P\{M \| \mathscr{G}\}$, где $\mathscr{G}$ есть σ-поле, порожденное этими величинами.

Пример 9.4. Рассмотрим марковский сдвиг (пример 3.1). Мера P определяется соотношением

$$P\{x_n = i_n, \ldots, x_{n+l} = i_{n+l}\} = p_{i_n} p_{i_n i_{n+1}} \ldots p_{i_{n+l-1} i_{n+l}}.$$

(В этом примере существенно именно это соотношение, а не сам сдвиг.) По формуле условной вероятности относительно конечного поля имеем

$$P\{x_0 = i \| x_{-n}, \ldots, x_{-1}\}_\omega = p_{x_{-1}(\omega), i} \text{ п. в.}^{1)}. \tag{9.8}$$

Но для того, чтобы доказать интуитивно очевидную формулу

$$P\{x_0 = i \| \ldots x_{-2}, x_{-1}\}_\omega = p_{x_{-1}(\omega), i} \text{ п. в.}, \tag{9.9}$$

потребуется следующая теорема, простым следствием которой эта формула и является.

Теорема 9.1. *Пусть $\mathscr{G}_0$ — (конечно аддитивное) поле, порождающее σ-поле $\mathscr{G}$. Интегрируемая функция f является вариантом функции $P\{M \| \mathscr{G}\}$, если она измерима относительно $\mathscr{G}$ и*

$$\int_G f\,dP = P(M \cap G) \tag{9.10}$$

для всех G из $\mathscr{G}_0$.

Доказательство. Каждая часть равенства (9.10) как функция множества G является мерой на $\mathscr{G}$. Так как функции совпадают на $\mathscr{G}_0$, то они должны совпадать и на $\mathscr{G}$.

[1]) Случайные величины x_n принимают не числовые значения; их значения лежат в конечном множестве ρ. Условные вероятности таких величин все еще определяются в терминах σ-полей, которые они порождают.

Пример 9.5. Пусть T — диадическое преобразование $T\omega=2\omega\,(\mathrm{mod}\,1)$ на единичном интервале с мерой Лебега $(\Omega, \mathscr{F}, P)$ (см. пример 1.6). В § 1 мы показали, что если M инвариантно и F лежит в поле $\mathscr{F}_0$, состоящем из конечных объединений непересекающихся диадических интервалов, то M и F независимы, так что $\int\limits_F P(M)\,dP=P(M\cap F)$. Так как $\mathscr{F}_0$ порождает $\mathscr{F}$, то из теоремы 9.1 следует $P\{M\|\mathscr{F}\}=P(M)$. Но (см. пример 9.1) $P\{M\|\mathscr{F}\}=I_M$ п. в. Поэтому $P(M)$ равна либо 0, либо 1, и мы получаем новое доказательство эргодичности преобразования T.

Следующий пример показывает, что интерпретация условной вероятности, использующая поведение Тихе, не проходит в некоторых патологических случаях.

Пример 9.6. Пусть $(\Omega, \mathscr{F}, P)$ — единичный интервал Ω с мерой Лебега P на σ-поле $\mathscr{F}$ борелевских множеств и $\mathscr{G}$ есть σ-подполе множеств, которые либо счетны, либо имеют счетные дополнения. Тогда функция, тождественно равная $P(M)$, является вариантом функции $P\{M\|\mathscr{G}\}$:

$$P\{M\|\mathscr{G}\}_\omega=P(M) \text{ п. в.} \tag{9.11}$$

Но так как $\mathscr{G}$ содержит все одноточечные множества, то знание того, какие элементы из $\mathscr{G}$ содержат ω, а какие нет, равносильно знанию самой точки ω. Таким образом, после получения сообщения Тихе экспериментатор знает, наступило событие $\omega\in M$ или нет, и мы должны иметь

$$P\{M\|\mathscr{G}\}_\omega=\begin{cases}1, & \text{если} \quad \omega\in M,\\ 0, & \text{если} \quad \omega\notin M.\end{cases} \tag{9.12}$$

Математическое определение приводит к (9.11), эвристическое — к (9.12). Разумеется, (9.11) правильно, а (9.12) — нет[1]).

Несмотря на то что наша интерпретация в некоторых случаях непригодна, она дает удобный способ представления условных вероятностей и не может привести к трудностям, так как доказательства на нее не опираются.

---

[1]) В этом частном случае можно избежать трудностей, допуская варианты, почти всюду равные некоторой измеримой относительно $\mathscr{G}$ функции. Более сложные примеры (те самые, которые демонстрируют возможное несуществование условных вероятностных мер, см. конец этого параграфа) показывают, что это не всегда помогает.

Свойства условных вероятностей

Для любого варианта условной вероятности имеем $\int\limits_G P\{M \| \mathscr{G}\}\, dP = P(M \cap G) \geqslant 0$ при $G \in \mathscr{G}$. Так как $P\{M \| \mathscr{G}\}$ измерима относительно $\mathscr{G}$, то она почти всюду неотрицательна. Аналогичное рассуждение показывает, что она почти всюду не превосходит 1:

$$(\mathrm{P}_1) \qquad 0 \leqslant P\{M \| \mathscr{G}\} \leqslant 1 \text{ п. в.}$$

Если $P(M) = 1$, то $\int\limits_G P\{M \| \mathscr{G}\}\, dP = P(G)$ для всякого $G \in \mathscr{G}$; следовательно,

$$(\mathrm{P}_2) \qquad P\{M \| \mathscr{G}\} = 1 \text{ п. в., если } P(M) = 1$$

Аналогично

$$(\mathrm{P}_2') \qquad P\{M \| \mathscr{G}\} = 0 \text{ п. в., если } P(M) = 0$$

Докажем, что

$$(\mathrm{P}_3) \quad P\{M_1 \cup M_2 \| \mathscr{G}\} = P\{M_1 \| \mathscr{G}\} + P\{M_2 \| \mathscr{G}\} \text{ п. в., если } M_1 \cap M_2 = \varnothing.$$

Мы должны показать, что правая часть равенства (P_3) удовлетворяет двум требованиям, налагаемым на варианты функции $P\{M_1 \cup M_2 \| \mathscr{G}\}$. Ясно, что она измерима относительно $\mathscr{G}$. Так как M_1 и M_2 не пересекаются, то для $G \in \mathscr{G}$

$$\begin{aligned}\int\limits_G (P\{M_1 \| \mathscr{G}\} + P\{M_2 \| \mathscr{G}\})\, dP &= \\ &= \int\limits_G P\{M_1 \| \mathscr{G}\}\, dP + \int\limits_G P\{M_2 \| \mathscr{G}\}\, dP = \\ &= P(M_1 \cap G) + P(M_2 \cap G) = \\ &= P((M_1 \cup M_2) \cap G),\end{aligned}$$

так что функциональное равенство выполнено.

Аналогичными рассуждениями можно доказать, что

$$(\mathrm{P}_3') \quad P\{M_1 - M_2 \| \mathscr{G}\} = P\{M_1 \| \mathscr{G}\} - P\{M_2 \| \mathscr{G}\} \text{ п. в., если } M_2 \subset M_1;$$

$$(\mathrm{P}_3'') \quad P\{M^c \| \mathscr{G}\} = 1 - P\{M \| \mathscr{G}\} \text{ п. в.};$$

$$(\mathrm{P}_3''') \quad |P\{M_1 \| \mathscr{G}\} - P\{M_2 \| \mathscr{G}\}| \leqslant P\{M_1 + M_2 \| \mathscr{G}\} \text{ п. в. } ^{1)}$$

---

[1]) Как обычно, + означает симметрическую разность.

Предположим теперь, что $\{M_n\}$ — невозрастающая последовательность множеств с пересечением M, это обозначается $M_n \downarrow M$. В силу (P_3') и (P_1) последовательность $P\{M_n \| \mathscr{G}\}_\omega$ п. в. не возрастает и, следовательно, имеет предел $f(\omega)$ п. в. Функция f измерима относительно $\mathscr{G}$; в силу (P_1) и сходимости ограниченной монотонной последовательности

$$\int_G f\,dP = \lim_{n\to\infty} \int_G P\{M_n \| \mathscr{G}\}\,dP = \lim_{n\to\infty} P(M_n \cap G) = P(M \cap G)$$

для любого $G \in \mathscr{G}$. Поэтому f есть вариант функции $P\{M \| \mathscr{G}\}$:

(P_4) $\qquad P\{M_n \| \mathscr{G}\} \downarrow P\{M \| \mathscr{G}\}$ п. в., если $M_n \downarrow M$.

Аналогично

(P_4') $\qquad P\{M_n \| \mathscr{G}\} \uparrow P\{M \| \mathscr{G}\}$ п. в., если $M_n \uparrow M$.

Наконец, в силу (P_3) и (P_4') имеем

(P_5) $\qquad P\left\{\bigcup_n M_n \| \mathscr{G}\right\} = \sum_n P\{M_n \| \mathscr{G}\}$ п. в.,

если $M_m \cap M_n = \varnothing \quad (m \neq n)$.

Пример 9.7. Из соотношений (9.8) и (9.9) для цепей Маркова следует, что

$$P\{x_0 = i \| \ldots, x_{-2}, x_{-1}\} = P\{x_0 = i \| x_{-1}\} \text{ п. в.}$$

Аналогичное рассуждение показывает, что если M — цилиндр, зависящий от координат с неотрицательными индексами, то

$$P\{M \| \ldots, x_{-2}, x_{-1}\} = P\{M \| x_{-1}\} \text{ п. в.} \tag{9.13}$$

Используем только что доказанное свойство, чтобы показать, что равенство (9.13) справедливо, если M — произвольное множество из σ-поля, порожденного $x_0, x_1, \ldots$. Для любого такого M и любого положительного ε существует цилиндр M_ε, содержащий лишь координаты с неотрицательными индексами и такой, что $P(M + M_\varepsilon) < \varepsilon$. В силу (P_3''') имеем

$$\Delta = |P\{M \| x_{-1}\} - P\{M_\varepsilon \| x_{-1}\}| \leqslant \{M + M_\varepsilon \| x_{-1}\} \text{ п. в.}$$

Так как $\int_\Omega P\{M + M_\varepsilon \| x_{-1}\}\,dP = P(M + M_\varepsilon) < \varepsilon$, то

$$P\{\Delta > \lambda\} \leqslant \frac{E\{\Delta\}}{\lambda} \leqslant \frac{\varepsilon}{\lambda}.$$

Аналогично $|P\{M \| \ldots, x_{-2}, x_{-1}\} - P\{M_\varepsilon \| \ldots, x_{-2}, x_{-1}\}|$ превосходит λ с вероятностью, не большей ε/λ. Поэтому левая

и правая части равенства (9.13) отличаются друг от друга более чем на 2λ с вероятностью, не превосходящей $2\varepsilon/\lambda$. Полагая $\varepsilon \to 0$, а затем $\lambda \to 0$, получаем (9.13).

Функции и меры

В этом параграфе принята „глобальная" точка зрения. Мы связываем с каждым фиксированным M из $\mathscr{F}$ некоторую функцию (точнее, класс функций) $P\{M \| \mathscr{G}\}$, определенную на всем пространстве Ω. Что будет, если мы изменим точку зрения, фиксируя ω и предполагая, что M пробегает $\mathscr{F}$? Получим ли мы вероятностную меру на $\mathscr{F}$? Если это так, то, разумеется, только что доказанные результаты (P_1), ..., (P_5) сводятся к стандартным фактам, относящимся к мерам.

Пусть $\mathscr{G}$ — конечное поле с атомами $G_1, \ldots, G_r$. Если $P(G_1) = 0$ и $P(G_i) > 0$ для остальных i, то вариантом функции $P\{M \| \mathscr{G}\}$ будет

$$P\{M \| \mathscr{G}\}_\omega = \begin{cases} \dfrac{P(M \cap G_i)}{P(G_i)}, & \text{если } \omega \in G_i, \quad i = 2, \ldots, r, \\ 0, & \text{если } \omega \in G_1. \end{cases}$$

При таком выборе варианта условной вероятности $P\{M \| \mathscr{G}\}_\omega$ как функция от M является вероятностной мерой на $\mathscr{F}$, если $\omega \in G_2 \cup \ldots \cup G_r$, и не является мерой, если $\omega \in G_1$. Мы выбрали „неправильный" вариант. Если взять, скажем, вариант

$$P\{M \| \mathscr{G}\}_\omega = \begin{cases} \dfrac{P(M \cap G_i)}{P(G_i)}, & \text{если } \omega \in G_i, \quad i = 2, \ldots, r, \\ P(M), & \text{если } \omega \in G_1, \end{cases}$$

то $P\{M \| \mathscr{G}\}_\omega$ есть вероятностная мера по M для каждого ω. Ясно, что такие варианты существуют, если $\mathscr{G}$ конечно.

Можно было бы подумать, что для любого σ-поля $\mathscr{G}$ варианты различных $P\{M \| \mathscr{G}\}$ могут быть выбраны так, что $P\{M \| \mathscr{G}\}_\omega$ являются вероятностными мерами по M для каждого[1] ω из Ω. Можно показать на примере, что это не так (см. Дуб [1]).

Предположим, что точка ω_0 обладает тем свойством, что $P(G) > 0$ для любого G из $\mathscr{G}$, содержащего ω_0. Это верно, например, если одноточечное множество $\{\omega_0\}$ принадлежит $\mathscr{F}$ и имеет положительную меру. Фиксируем любые варианты функции $P\{M \| \mathscr{G}\}$ для всех M из $\mathscr{F}$. Для любого M мно-

---

[1]) Если это можно сделать для почти всех ω, то ясно, что это можно сделать и для всех ω.

жество $\{\omega: P\{M \| \mathscr{G}\}_\omega < 0\}$ лежит в $\mathscr{G}$ и в силу (P_1) имеет меру 0, а следовательно, не может содержать ω_0. Таким образом, $P\{M \| \mathscr{G}\}_{\omega_0} \geqslant 0$. Аналогично $P\{\Omega \| \mathscr{G}\}_{\omega_0} = 1$, и если M_n не пересекаются, то $P\left\{\bigcup_n M_n \| \mathscr{G}\right\}_{\omega_0} = \sum_n P\{M_n \| \mathscr{G}\}_{\omega_0}$. Следовательно, $P\{M \| \mathscr{G}\}_{\omega_0}$ есть вероятностная мера, если M пробегает $\mathscr{F}$.

Таким образом, условные вероятности ведут себя должным образом в точках, имеющих положительную вероятность. То, что они могут плохо вести себя в точках вероятности 0, не вызывает затруднений, поскольку отдельные такие точки не влияют на вероятности множеств. Разумеется, *множество* точек, каждая из которых имеет меру 0, оказывает влияние, но здесь мы возвращаемся к глобальной точке зрения, из которой мы исходили [1]).

З а м е ч а н и е. Колмогоровым [1] было введено впервые общее понятие условных вероятностей. Приведенное изложение следует во всем Дубу [1].

10. УСЛОВНЫЕ МАТЕМАТИЧЕСКИЕ ОЖИДАНИЯ

Определение

Пусть x — интегрируемая случайная величина на $(\Omega, \mathscr{F}, P)$ и $\mathscr{G}$ есть σ-подполе поля $\mathscr{F}$. Определим функцию φ множества на $\mathscr{G}$ равенством

$$\varphi(G) = \int_G x\, dP, \qquad G \in \mathscr{G}.$$

Тогда φ — конечнозначная (даже ограниченная), вполне аддитивная функция. Более того, она абсолютно непрерывна относительно P (рассматриваемой на $\mathscr{G}$). В силу теоремы Радона — Никодима существует функция $E\{x \| \mathscr{G}\}$, интегрируемая, измеримая относительно $\mathscr{G}$ и удовлетворяющая равенству $\varphi(G) = \int_G E\{x \| \mathscr{G}\}\, dP$ для всех $G \in \mathscr{G}$. Обозначим через $E\{x \| \mathscr{G}\}_\omega$ значение функции $E\{x \| \mathscr{G}\}$ в точке ω.

---

[1]) Желание сохранить локальную точку зрения, наложив на пространство с мерой подходящие ограничения, приводит к понятиям пространства Лебега и измеримого разбиения; см. Рохлин [7]. Для эргодической теории эти понятия очень полезны. — *Прим. ред.*

Будем называть эту функцию условным математическим ожиданием (или ожидаемым значением) величины x относительно $\mathscr{G}$. Она определяется требованиями, чтобы a) $E\{x\,\|\,\mathscr{G}\}$ была интегрируемой и измеримой относительно $\mathscr{G}$ и b) $E\{x\,\|\,\mathscr{G}\}$ удовлетворяла функциональному уравнению

$$\int_G E\{x\,\|\,\mathscr{G}\}\,dP = \int_G x\,dP, \qquad G \in \mathscr{G}. \tag{10.1}$$

Вообще говоря, существует много таких функций; их называют вариантами условного математического ожидания. Любые два варианта совпадают почти всюду.

Если $\mathscr{G} = 2$, то функция, тождественно равная $E\{x\}$, является вариантом функции $E\{x\,\|\,\mathscr{G}\}$. Если $\mathscr{G} = \mathscr{F}$, то x само есть вариант функции $E\{x\,\|\,\mathscr{G}\}$. Для любого $\mathscr{G}$ функция, тождественно равная $E\{x\}$, удовлетворяет приведенному выше условию (a) (которое становится более ограничительным с уменьшением $\mathscr{G}$), а само x удовлетворяет условию (b) (которое становится более ограничительным с расширением $\mathscr{G}$). Эти два условия работают в противоположных направлениях, и между ними находится класс вариантов функции $E\{x\,\|\,\mathscr{G}\}$.

Мы интерпретируем $E\{x\,\|\,\mathscr{G}\}_\omega$ как новое ожидаемое значение случайной величины x для экспериментатора, когда Тихе сообщает ему, какие из множеств в $\mathscr{G}$ содержат выбранную ею точку ω, а какие нет. Условия (a) и (b) соответствуют аналогичным в определении условной вероятности. Согласно первому, экспериментатор может вычислить значение $E\{x\,\|\,\mathscr{G}\}$ после получения сообщения Тихе. Если после этого сообщения она предлагает ему выплачивать x драхм в ответ на некоторую предварительную ставку, то справедливая начальная ставка равна $E\{x\,\|\,\mathscr{G}\}$ драхм при выполнении интуитивных требований, связанных с условным математическим ожиданием. Если стратегия экспериментатора заключается в том, что он принимает пари в случае наступления G и отвергает его в противном случае (здесь $G \in \mathscr{G}$), то математическое ожидание его выигрыша равно $\int_G (x - E\{x\,\|\,\mathscr{G}\})\,dP$. Соотношение (10.1) требует, чтобы такая стратегия не давала преимуществ ни экспериментатору, ни богине.

Если $\mathscr{G}$ — конечное поле, то $E\{x\,\|\,\mathscr{G}\}_\omega$ должно равняться $\int_G x\,dP/P(G_i)$ для ω из атома G_i, имеющего положительную вероятность, в то время как $E\{x\,\|\,\mathscr{G}\}$ может принимать

любое постоянное значение на атоме с нулевой вероятностью. Это связывает наше общее определение с классическим понятием условного математического ожидания при заданном множестве G:

$$E\{x \mid G\} = \frac{1}{P(G)} \int_G x \, dP \qquad (P(G) > 0).$$

Заметим, что для множества $M \in \mathscr{F}$ определяющие свойства функции $E\{I_M \| \mathscr{G}\}$ и $P\{M \| \mathscr{G}\}$ совпадают. Таким образом,

$$E\{I_M \| \mathscr{G}\} = P\{M \| \mathscr{G}\} \quad \text{п. в.},$$

что обобщает соотношение $E\{I_M\} = P(M)$.

Пример 10.1. Если $x = \sum_i a_i I_{A_i}$ — простая функция, то

$$E\{x \| \mathscr{G}\} = \sum_i a_i P\{A_i \| \mathscr{G}\} \quad \text{п. в.}$$

Пример 10.2. Исследуем предельную функцию $\hat{f}$ в эргодической теореме. Если f интегрируема, то в соответствии с этой теоремой

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k \omega) = \hat{f}(\omega) \quad \text{п. в.} \tag{10.2}$$

Для каждого множества A произведение $I_A f$ имеет свою „функцию с крышечкой" $(I_A f)^\wedge$, к которой оно сходится в среднем:

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} I_A(T^k \omega) f(T^k \omega) = (I_A f)^\wedge(\omega) \quad \text{п. в.} \tag{10.3}$$

Если A инвариантно, так что $I_A(T^k \omega) = I_A(\omega)$ п. в., то из (10.2) и (10.3) следует $I_A(\omega) \hat{f}(\omega) = (I_A f)^\wedge(\omega)$ п. в. Так как $(I_A f)^\wedge$ и $I_A f$ имеют одинаковые математические ожидания, то

$$\int I_A \hat{f} \, dP = \int (I_A f)^\wedge \, dP = \int I_A f \, dP$$

или

$$\int_A \hat{f} \, dP = \int_A f \, dP. \tag{10.4}$$

Очевидно, что класс $\mathscr{I}$ множеств, инвариантных относительно T, образует σ-подполе поля $\mathscr{F}$ и функция $\hat{f}$, будучи инвариантной, измерима относительно $\mathscr{I}$. Так как (10.4)

выполнено для всех $A \in \mathcal{G}$, то $\bar{f}$ есть вариант функции $E\{f \| \mathcal{G}\}$.

Пример 10.3. Легко показать с помощью замены переменной, что если $y(\omega) = x(T\omega)$, то

$$E\left\{y \| T^{-1}\mathcal{G}\right\}_{\omega} = E\{x \| \mathcal{G}\}_{T\omega} \quad \text{п. в.} \tag{10.5}$$

В частности,

$$P\left\{T^{-1}M \| T^{-1}\mathcal{G}\right\}_{\omega} = P\{M \| \mathcal{G}\}_{T\omega} \quad \text{п. в.} \tag{10.6}$$

Основные свойства

Большинство примеров и результатов предыдущего параграфа переносятся непосредственно. Например, теорема 9.1 имеет очевидный аналог. Кроме того, те же методы, которые были использованы для доказательства (P_1), (P_2) и (P_3), достаточны для установления следующих трех результатов. Мы всюду предполагаем, что величины x, y и т. д. интегрируемы.

(E_1) Если $x = a$ п. в. (a — постоянная), то $E\{x \| \mathcal{G}\} = a$ п. в.

(E_2) Если $x \leqslant y$ п. в., то $E\{x \| \mathcal{G}\} \leqslant E\{y \| \mathcal{G}\}$ п. в.

(E_3) Если a и b — постоянные, то $E\{ax + by \| \mathcal{G}\} = aE\{x \| \mathcal{G}\} + bE\{y \| \mathcal{G}\}$ п. в.

Из этих утверждений вытекает очевидное следствие, что если $x = y$ п. в., то $E\{x \| \mathcal{G}\} = E\{y \| \mathcal{G}\}$ п. в.

Так как $-|x| \leqslant x \leqslant |x|$, то в силу ($E_2$) и ($E_3$)

$$-E\{|x| \| \mathcal{G}\} \leqslant E\{x \| \mathcal{G}\} \leqslant E\{|x| \| \mathcal{G}\} \quad \text{п. в.}$$

и, следовательно,

(E_4) $$|E\{x \| \mathcal{G}\}| \leqslant E\{|x| \| \mathcal{G}\} \quad \text{п. в.}$$

Нам понадобится следующее обобщение теоремы Лебега о переходе к пределу под знаком интеграла:

(E_5) Если $\lim x_n = x$ п. в. и $|x_n| \leqslant y$ п. в.,

где y интегрируема, то $\lim\limits_n E\{x_n \| \mathcal{G}\} = E\{x \| \mathcal{G}\}$ п. в. Для доказательства рассмотрим $z_n = \sup\limits_{k \geqslant n} |x_k - x|$. Имеем $z_n \downarrow 0$ п. в. и $|E\{x_n \| \mathcal{G}\} - E\{x \| \mathcal{G}\}| \leqslant E\{z_n \| \mathcal{G}\}$ п. в. в силу (E_3) и (E_4). Достаточно поэтому показать, что $E\{z_n \| \mathcal{G}\} \downarrow 0$ п. в. В силу (E_2) последовательность $E\{z_n \| \mathcal{G}\}$ п. в. не возрастает и, следовательно, имеет п. в. предел g. Нужно доказать, что

$g=0$ п. в. или, поскольку g неотрицательна, что $E\{g\}=0$. Но в силу теоремы о переходе к пределу под знаком интеграла, примененной к величинам $z_n(|z_n|\leqslant 2y)$,

$$E\{g\}\leqslant \int E\{z_n\,\|\,\mathscr{G}\}\,dP=\int z_n\,dP\to 0.$$

Если x измерима относительно $\mathscr{G}$, то, разумеется, $E\{x\,\|\,\mathscr{G}\}=x$ п. в. Постоянно приходится пользоваться следующим обобщением.

Теорема 10.1. *Если x измерима относительно $\mathscr{G}$ и если y и xy интегрируемы, то*

$$E\{xy\,\|\,\mathscr{G}\}=xE\{y\,\|\,\mathscr{G}\}\quad \text{п. в.} \tag{10.7}$$

Доказательство. Предположим сначала, что $x=I_{G_0}$, где $G_0\in\mathscr{G}$. Нужно показать, что $I_{G_0}E\{y\,\|\,\mathscr{G}\}$ есть вариант функции $E\{I_{G_0}y\,\|\,\mathscr{G}\}$. Так как $I_{G_0}E\{y\,\|\,\mathscr{G}\}$, очевидно, измерима относительно $\mathscr{G}$, то нужно только проверить функциональное равенство

$$\int\limits_G I_{G_0}E\{y\,\|\,\mathscr{G}\}\,dP=\int\limits_G I_{G_0}y\,dP,\qquad G\in\mathscr{G}.$$

Но это сводится к равенству

$$\int\limits_{G\cap G_0} E\{y\,\|\,\mathscr{G}\}\,dP=\int\limits_{G\cap G_0} y\,dP,$$

которое выполняется в силу определения $E\{y\,\|\,\mathscr{G}\}$. Таким образом, (10.7) выполнено, если x — характеристическая функция множества из $\mathscr{G}$. Далее, для любой функции x, измеримой относительно $\mathscr{G}$, существуют ступенчатые функции x_n, измеримые относительно $\mathscr{G}$ и такие, что $|x_n|\leqslant|x|$ и $\lim\limits_n x_n=x$. Так как $|x_ny|\leqslant|xy|$ и функция $|xy|$ интегрируема, то из (E_5) вытекает $\lim\limits_n E\{x_ny\,\|\,\mathscr{G}\}=E\{xy\,\|\,\mathscr{G}\}$ п. в. Но $E\{x_ny\,\|\,\mathscr{G}\}=x_nE\{y\,\|\,\mathscr{G}\}$ в силу уже доказанного и, разумеется, $\lim\limits_n x_nE\{y\,\|\,\mathscr{G}\}=xE\{y\,\|\,\mathscr{G}\}$. Поэтому $E\{xy\,\|\,\mathscr{G}\}=xE\{y\,\|\,\mathscr{G}\}$. Заметим, что мы не предполагали функцию x интегрируемой.

Повторные условные математические ожидания

Взятие условного математического ожидания можно рассматривать как операцию осреднения, или сглаживающую операцию. Переходя от x к $E\{x\,\|\,\mathscr{G}\}$, мы делаем x „более

близкой к константе". Это можно видеть графически на единичном интервале с мерой Лебега, если $\mathscr{G}$ конечно и имеет своими атомами подинтервалы. Естественно ожидать поэтому, что осреднение x относительно σ-поля $\mathscr{G}_2$, а затем осреднение результата относительно более „грубого" (меньшего) σ-поля $\mathscr{G}_1$ должно привести к тому же результату, что и приведенное сразу осреднение относительно $\mathscr{G}_1$.

Теорема 10.2. *Если x интегрируема и σ-поля $\mathscr{G}_1$ и $\mathscr{G}_2$ удовлетворяют соотношению $\mathscr{G}_1 \subset \mathscr{G}_2$, то*

$$E\{E\{x \| \mathscr{G}_2\} \| \mathscr{G}_1\} = E\{x \| \mathscr{G}_1\} \text{ п. в.} \tag{10.8}$$

Доказательство. Докажем, что $E\{E\{x \| \mathscr{G}_2\} \| \mathscr{G}_1\}$ есть вариант функции $E\{x \| \mathscr{G}_1\}$. Разумеется, $E\{E\{x \| \mathscr{G}_2\} \| \mathscr{G}_1\}$ измерима относительно $\mathscr{G}_1$. Если $G \in \mathscr{G}_1$, то

$$\int_G E\{E\{x \| \mathscr{G}_2\} \| \mathscr{G}_1\}\, dP = \int_G E\{x \| \mathscr{G}_2\}\, dP;$$

но так как G лежит также и в $\mathscr{G}_2$, правая часть этого равенства равна $\int_G x\, dP$. Таким образом, $E\{E\{x \| \mathscr{G}_2\} \| \mathscr{G}_1\}$ удовлетворяет функциональному уравнению для вариантов функции $E\{x \| \mathscr{G}_1\}$.

Мы получим второе доказательство равенства (10.8), установив, что $E\{x \| \mathscr{G}_1\}$ есть вариант функции $E\{E\{x \| \mathscr{G}_2\} \| \mathscr{G}_1\}$. Так как $E\{x \| \mathscr{G}_1\}$ измерима относительно $\mathscr{G}_1$, то нужно только проверить функциональное равенство

$$\int_G E\{x \| \mathscr{G}_1\}\, dP = \int_G E\{x \| \mathscr{G}_2\}\, dP, \quad G \in \mathscr{G}_1.$$

Но если $G \in \mathscr{G}_1$, то $G \in \mathscr{G}_2$ и обе части равенства совпадают с $\int_G x\, dP$.

Если $\mathscr{G}_2 = \mathscr{F}$, то $E\{x \| \mathscr{G}_2\} = x$ п. в., так что (10.8) тривиально. Если $\mathscr{G}_1 = 2$, то (10.8) принимает вид $E\{E\{x \| \mathscr{G}_2\}\} = E\{x\}$ — частный случай формулы (10.1).

Если $\mathscr{G}_1 \subset \mathscr{G}_2$, то случайная величина $E\{x \| \mathscr{G}_1\}$, будучи измеримой относительно $\mathscr{G}_1$, измерима и относительно $\mathscr{G}_2$; она не изменяется при взятии ее математического ожидания относительно $\mathscr{G}_2$:

$$E\{E\{x \| \mathscr{G}_1\} \| \mathscr{G}_2\} = E\{x \| \mathscr{G}_1\} \text{ п. в.}$$

Но это дополнение к (10.8) бесполезно.

Неравенство Йенсена

Для обычных интегралов и математических ожиданий неравенство Йенсена состоит в том, что если функция φ вещественна и выпукла [1]), то

$$\varphi(E\{x\}) \leqslant E\{\varphi(x)\} \tag{10.9}$$

в предположении, что математическое ожидание существует. Обобщение на случай условных математических ожиданий таково: если x и $\varphi(x)$ интегрируемы, то

$$\varphi(E\{x \| \mathscr{G}\}) \leqslant E\{\varphi(x) \| \mathscr{G}\} \text{ п. в.} \tag{10.10}$$

Если $\mathscr{G} = 2$, то (10.10) сводится к (10.9); если $\varphi(t) = |t|$, оно сводится к (E_4).

Мы докажем (10.10) только для одного случая, который нам нужен, а именно когда x принимает значения из конечного отрезка $[a, b]$, а φ определена на этом отрезке. В этом случае φ непрерывна и ограничена. Если $A_1, \ldots, A_r$ есть $\mathscr{F}$-разбиение пространства Ω, то $\sum P\{A_i \| \mathscr{G}\}_\omega = 1$ для почти всех ω. Для любого такого ω и любых чисел $\alpha_1, \ldots, \alpha_r$ из $[a, b]$ имеем

$$\varphi\left(\sum \alpha_i P\{A_i \| \mathscr{G}\}_\omega\right) \leqslant \sum_i \varphi(\alpha_i) P\{A_i \| \mathscr{G}\}_\omega$$

в силу основного свойства выпуклых функций. Поэтому (10.10) выполняется для простых случайных величин. Но любая случайная величина x со значениями из $[a, b]$ является пределом простых случайных величин x_n со значениями из $[a, b]$. Так как последовательности $\{x_n\}$ и $\{\varphi(x_n)\}$ равномерно ограничены, то (10.10) вытекает из (E_5).

Одна специальная формула

В дальнейшем нам понадобится тот факт, что если $\mathscr{A}$ — конечное поле с атомами $A_1, \ldots, A_r$, то

$$P\{M \| \mathscr{A} \vee \mathscr{G}\} = \sum_{i=1}^{r} I_{A_i} \frac{P\{M \cap A_i \| \mathscr{G}\}}{P\{A_i \| \mathscr{G}\}} \text{ п. в.} \tag{10.11}$$

Для почти всех ω, для которых знаменатель в правой части обращается в нуль, числитель также обращается в нуль, и мы принимаем отношение равным 0. Если $\mathscr{G} = 2$, то (10.11)

[1]) Изящное изложение вопросов выпуклости можно найти в гл. 1 книги Зигмунда [1].

обращается в исходную формулу для условных вероятностей относительного конечного поля $\mathcal{A}$.

Покажем, что сумма в правой части равенства (10.11) является вариантом функции $P\{M \| \mathcal{A} \vee \mathcal{G}\}$. Так как эта сумма измерима относительно $\mathcal{A} \vee \mathcal{G}$, нужно только проверить выполнение функционального равенства. Но $\mathcal{A} \vee \mathcal{G}$ состоит из конечных объединений непересекающихся множеств вида $A_j \cap G$, где $G \in \mathcal{G}$, следовательно, достаточно проверить, что

$$\int\limits_{A_j \cap G} \sum_i I_{A_i} g_i \, dP = P(M \cap A_j \cap G), \tag{10.12}$$

где $g_i = P\{M \cap A_i \| \mathcal{G}\} / P\{A_i \| \mathcal{G}\}$. Так как A_i не пересекаются и g_j измеримы относительно $\mathcal{G}$, то левая часть равенства (10.12) равна

$$\int\limits_{A_j \cap G} I_{A_j} g_j \, dP = \int\limits_G I_{A_j} g_j \, dP = \int\limits_G E\{I_{A_j} g_j \| \mathcal{G}\} \, dP =$$

$$= \int\limits_G g_j P\{A_j \| \mathcal{G}\} \, dP = \int\limits_G P\{M \cap A_j \| \mathcal{G}\} \, dP = P(M \cap A_j \cap G).$$

Пример 10.4. Пусть A — такое множество, для которого $P(A) > 0$. Определим меру P_1 соотношением

$$P_1(B) = P(B \mid A), \quad B \in \mathcal{F}. \tag{10.13}$$

Докажем одну формулу, связанную с (10.11):

$$P_1\{M \| \mathcal{G}\} = \frac{P\{M \cap A \| \mathcal{G}\}}{P\{A \| \mathcal{G}\}} \text{ п. в. } P_1 \tag{10.14}$$

(равенство имеет место с точностью до множества P_1-меры 0). Заметим, что в силу абсолютной непрерывности P_1 относительно P обе части равенства (10.14) определены с точностью до множеств P_1-меры 0. Правая часть равенства (10.14) измерима относительно $\mathcal{G}$; нужно показать, что она удовлетворяет равенству

$$\int\limits_G \frac{P\{M \cap A \| \mathcal{G}\}}{P\{A \| \mathcal{G}\}} \, dP_1 = P_1(M \cap G), \quad G \in \mathcal{G},$$

которое эквивалентно

$$\int\limits_{G \cap A} \frac{P\{M \cap A \| \mathcal{G}\}}{P\{A \| \mathcal{G}\}} \, dP = P(M \cap A \cap G), \quad G \in \mathcal{G}.$$

Но если $G \in \mathscr{G}$, то левая часть этого соотношения равна

$$\int\limits_G I_A \frac{P\{M \cap A \| \mathscr{G}\}}{P\{A \| \mathscr{G}\}} dP = \int\limits_G E\left\{I_A \frac{P\{M \cap A \| \mathscr{G}\}}{P\{A \| \mathscr{G}\}} \middle\| \mathscr{G}\right\} dP =$$

$$= \int\limits_G \frac{P\{M \cap A \| \mathscr{G}\}}{P\{A \| \mathscr{G}\}} E\{I_A \| \mathscr{G}\} dP = \int\limits_G P\{M \cap A \| \mathscr{G}\} dP = P(M \cap A \cap G),$$

что доказывает (10.14). Если A оказывается лежащим в $\mathscr{G}$, то (10.14) принимает вид

$$P_1\{M \| \mathscr{G}\} = P\{M \| \mathscr{G}\} \text{ п. в. } P_1. \tag{10.15}$$

Пример 10.5. Если T — неэргодический сдвиг, то он не имеет своим респектом некоторый цилиндр C. Так как предельная функция

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} I_C(T^k\omega) = \hat{I}_C(\omega) \tag{10.16}$$

не является тогда константой почти всюду, то существует такое a, что инвариантное множество $A_0 = \{\omega : \hat{I}_C(\omega) < a\}$ удовлетворяет условию $0 < P(A_0) < 1$. Так как C — цилиндр, то A_0 лежит в σ-поле, порожденном $x_n, x_{n+1}, \ldots$ для любого n, что следует из (10.16). Обращая временну́ю шкалу (рассматривая T^{-1}), мы видим, что существует также нетривиальное инвариантное множество A, которое для любого n принадлежит σ-полю, порожденному $\ldots, x_{n-1}, x_n$.

Определим P_1 равенством (10.13); в силу (10.15) имеем

$$P_1\{M \| \ldots, x_{n-1}, x_n\} = P\{M \| \ldots, x_{n-1}, x_n\} \text{ п. в. } P_1$$

для каждого $n = 0, \pm 1, \ldots$ и для каждого M из $\mathscr{F}$. В частности,

$$P_1\{x_{n+1} = i \| \ldots, x_{n-1}, x_n\} = P\{x_{n+1} = i \| \ldots, x_{n-1}, x_n\} \text{ п. в. } P_1. \tag{10.17}$$

Поэтому если T неэргодично относительно P, то существует вероятностная мера P_1 на $\mathscr{F}$, которая отлична от P, абсолютно непрерывна относительно нее и имеет ту же структуру, что и P [в смысле (10.17)].

Предположим теперь, что T — неэргодический марковский сдвиг относительно P с матрицей вероятностей перехода Π и стационарным распределением p. Из (10.17) следует, что T относительно P_1 также является марковским сдвигом с матрицей вероятностей перехода Π. Если P_1 отлична от P, то сдвиг должен иметь другое стационарное распределение,

Таким образом, T эргодично, если существует только одно стационарное распределение для П; это уже было получено в § 3 другим способом.

Замечание. Дальнейшие сведения относительно условных математических ожиданий см. Дуб [1].

11. ТЕОРЕМА СХОДИМОСТИ

Можно задать вопрос, будет ли функция $E\{x\|\mathcal{G}\}$ при изменении $\mathcal{G}$ непрерывна в каком-нибудь смысле. Пусть $\mathcal{G}_1 \subset \mathcal{G}_2 \subset \ldots$ и $\mathcal{G} = \bigvee_{n=1}^{\infty} \mathcal{G}_n$ (мы будем обозначать это записью $\mathcal{G}_n \uparrow \mathcal{G}$). В этом параграфе мы покажем, что тогда

$$\lim_{n\to\infty} E\{x\|\mathcal{G}_n\} = E\{x\|\mathcal{G}\} \text{ п. в.}$$

Теорема

Докажем сначала результат, который служит той же цели, что и максимальная эргодическая теорема.

Теорема 11.1. *Предположим, что функция x интегрируема и $\mathcal{G}_1 \subset \mathcal{G}_2 \subset \ldots$. Тогда для всех положительных λ*

$$P\left\{\sup_{n\geqslant 1} E\{|x|\|\mathcal{G}_n\} > \lambda\right\} \leqslant \frac{1}{\lambda} E\{|x|\}. \tag{11.1}$$

Доказательство. Достаточно показать, что для каждого n справедливо неравенство

$$P\left\{\max_{1\leqslant k\leqslant n} E\{|x|\|\mathcal{G}_k\} > \lambda\right\} \leqslant \frac{1}{\lambda} E\{|x|\}. \tag{11.2}$$

Для $n=1$ это — неравенство чебышевского типа, так как $E\{|x|\|\mathcal{G}_1\}$ есть случайная величина с математическим ожиданием $E\{|x|\}$.

Чтобы перейти к общему случаю, рассмотрим множество M_k ($k\leqslant n$) точек ω, для которых $E\{|x|\|\mathcal{G}_i\} \leqslant \lambda$ при $i<k$ и $E\{|x|\|\mathcal{G}_k\} > \lambda$. Тогда $M_1, \ldots, M_n$ — непересекающиеся множества, объединение которых M есть множество, фигурирующее в (11.2). Если $i\leqslant k$, то функция $E\{|x|\|\mathcal{G}_i\}$, будучи измеримой относительно $\mathcal{G}_i$, измерима также относительно $\mathcal{G}_k$ (здесь мы используем предположение $\mathcal{G}_1 \subset \mathcal{G}_2 \subset \ldots$). Так как M_k определено в терминах $E\{|x|\|\mathcal{G}_i\}$, где $i\leqslant k$, то оно

лежит в $\mathscr{G}_k$. Следовательно,

$$\lambda P(M) = \sum_{k=1}^{n} \int_{M_k} \lambda \, dP \leqslant \sum_{k=1}^{n} \int_{M_k} E\{|x| \,\|\, \mathscr{G}_k\} \, dP = \sum_{k=1}^{n} \int_{M_k} |x| \, dP \leqslant E\{|x|\}.$$

Теорема доказана [1]).

В силу непрерывности по λ можно заменить знак $>$ в (11.1) знаком $\geqslant$. Можно также заменить $E\{|x| \,\|\, \mathscr{G}_n\}$ на $|E\{x \,\|\, \mathscr{G}_n\}|$ или на $E\{x \,\|\, \mathscr{G}_n\}$, поскольку $E\{x \,\|\, \mathscr{G}_n\} \leqslant |E\{x \,\|\, \mathscr{G}_n\}| \leqslant \leqslant E\{|x| \,\|\, \mathscr{G}_n\}$.

Теорема 11.2. *Если $\mathscr{G}_n \uparrow \mathscr{G}$, то для любой интегрируемой функции x*

$$\lim_{n \to \infty} E\{x \,\|\, \mathscr{G}_n\} = E\{x \,\|\, \mathscr{G}\} \text{ п. в.} \tag{11.3}$$

и для любого $M \in \mathscr{F}$

$$\lim_{n \to \infty} P\{M \,\|\, \mathscr{G}_n\} = P\{M \,\|\, \mathscr{G}\} \text{ п. в.} \tag{11.4}$$

Доказательство. Равенство (11.4) является, конечно, частным случаем равенства (11.3). Далее, если (11.3) выполнено для любой интегрируемой и измеримой относительно $\mathscr{G}$ функции x, то, даже когда она неизмерима относительно $\mathscr{G}$, это равенство имеет место, если заменить x на $E\{x \,\|\, \mathscr{G}\}$:

$$\lim_{n \to \infty} E\{E\{x \,\|\, \mathscr{G}\} \,\|\, \mathscr{G}_n\} = E\{E\{x \,\|\, \mathscr{G}\} \,\|\, \mathscr{G}\} \text{ п. в.} \tag{11.5}$$

Но (11.5) сводится к (11.3) по теореме 10.2. Следовательно, можно считать x измеримой относительно $\mathscr{G}$, и тогда (11.3) принимает вид [2])

$$\lim_{n \to \infty} E\{x \,\|\, \mathscr{G}_n\} = x \text{ п. в.} \tag{11.6}$$

Если x измерима относительно какого-нибудь $\mathscr{G}_k$, то $E\{x \,\|\, \mathscr{G}_n\} = x$ п. в. для $n \geqslant k$, и мы непосредственно получаем (11.6). Мы докажем (11.6) для произвольной x,

---

[1]) Неравенство (11.1) и похожее на него неравенство Колмогорова для сумм независимых случайных величин являются частными случаями мартингального неравенства. См. Дуб [1].

[2]) В дальнейшем нам понадобится только (11.4). Однако только что предпринятый шаг приводит нас от (11.4) к (11.6) при x, измеримой относительно $\mathscr{G}$ и, в этом случае, ограниченной. Другими словами, используя применяемую технику, доказать (11.4) не легче, чем (11.3).

измеримой относительно $\mathscr{G}$, аппроксимируя ее такими случайными величинами. Для заданного ε выберем такую случайную величину x_ε, измеримую относительно некоторого $\mathscr{G}_k$, что [1])

$$E\{|x-x_\varepsilon|\}<\varepsilon.$$

Имеем

$$|E\{x\|\mathscr{G}_m\}-E\{x\|\mathscr{G}_n\}|\leqslant$$
$$\leqslant|E\{x_\varepsilon\|\mathscr{G}_m\}-E\{x_\varepsilon\|\mathscr{G}_n\}|+2\sup_{k\geqslant 1}E\{|x-x_\varepsilon|\|\mathscr{G}_k\}. \quad (11.7)$$

Так как $E\{x_\varepsilon\|\mathscr{G}_n\}\to x_\varepsilon$ п. в., то первое слагаемое в правой части неравенства (11.7) стремится к нулю при $m, n\to\infty$. Следовательно, в силу теоремы 11.1

$$P\left\{\overline{\lim_{m,n\to\infty}}|E\{x\|\mathscr{G}_m\}-E\{x\|\mathscr{G}_n\}|>\lambda\right\}\leqslant\frac{2}{\lambda}E\{|x-x_\varepsilon|\}<\frac{2\varepsilon}{\lambda}.$$

Полагая $\varepsilon\to 0$, видим, что вероятность в левой части неравенства равна 0. Таким образом, последовательность $E\{x\|\mathscr{G}_n\}$ почти всюду фундаментальна [2]).

Следовательно, предел $y=\lim\limits_n E\{x\|\mathscr{G}_n\}$ существует почти всюду. Нам нужно идентифицировать y с x. Предполагая, что y интегрируема и интегрирование можно производить под знаком предела, получаем

$$\int_A y\,dP=\lim_{n\to\infty}\int_A E\{x\|\mathscr{G}_n\}\,dP. \quad (11.8)$$

Если $A\in\mathscr{G}_k$, то $\int_A E\{x\|\mathscr{G}_n\}\,dP=\int_A x\,dP$ для всех $n\geqslant k$, так что

$$\int_A y\,dP=\int_A x\,dP.$$

Но если это равенство имеет место для всех A из $\bigcup_k \mathscr{G}_k$, то оно имеет место и для всех A из $\mathscr{G}$, так что в силу измеримости x и y относительно $\mathscr{G}$ имеем $y=x$ п. в.

---

[1]) Это возможно, поскольку x может быть аппроксимирована (в среднем) ступенчатой функцией, измеримой относительно $\mathscr{G}$, и каждое множество, связанное с этой ступенчатой функцией, может быть аппроксимировано множеством из поля $\bigcup_n \mathscr{G}_n$.

[2]) Это рассуждение соответствует первому доказательству эргодической теоремы в § 2.

Следовательно, нужно доказать только интегрируемость y и оправдать (11.8) [1]). Так как в силу леммы Фату

$$\int |y|\,dP \leqslant \varliminf_{n\to\infty} \int |E\{x\,\|\,\mathscr{G}_n\}|\,dP \leqslant \varliminf_{n\to\infty} \int E\{|x|\,\|\,\mathscr{G}_n\}\,dP \leqslant E\{|x|\},$$

то y интегрируема. Чтобы установить (11.8), достаточно доказать, что

$$\lim_{n\to\infty} \int |E\{x\,\|\,\mathscr{G}_n\} - y|\,dP = 0. \tag{11.9}$$

Если $x_n = E\{x\,\|\,\mathscr{G}_n\}$, то

$$\int |x_n - y|\,dP \leqslant \int\limits_{\{|x_n| \leqslant \lambda\}} |x_n - y|\,dP + \int\limits_{\{|x_n| > \lambda\}} E\{|x|\,\|\,\mathscr{G}_n\}\,dP + \int\limits_{\{|x_n| > \lambda\}} |y|\,dP.$$

Так как множество $\{|x_n| > \lambda\}$ лежит в $\mathscr{G}_n$, то второе слагаемое в правой части последнего соотношения равно $\int\limits_{\{|x_n| > \lambda\}} |x|\,dP$; полагая $N_\lambda = \left\{\sup\limits_n |x_n| > \lambda\right\}$, получаем

$$\int |x_n - y|\,dP \leqslant \int\limits_{\{|x_n| \leqslant \lambda\}} |x_n - y|\,dP + \int\limits_{N_\lambda} (|x| + |y|)\,dP.$$

Для фиксированного λ первое слагаемое в правой части при $n \to \infty$ стремится к 0 в силу теоремы о переходе к пределу под знаком интеграла. Так как $P(N_\lambda) \to 0$ при $\lambda \to \infty$ в силу теоремы 11.1, то получаем (11.9) [2]).

Примеры

Пример 11.1. Если $A \in \bigvee\limits_{n=1}^{\infty} \mathscr{A}_n$, где $\mathscr{A}_n$ — конечные поля, для которых $\mathscr{A}_1 \subset \mathscr{A}_2 \subset \ldots$, то

$$p\lim_{n\to\infty} P\{A\,\|\,\mathscr{A}_n\} = I_A \tag{11.10}$$

---

[1]) Для ограниченной x это очевидно, поэтому (11.4) уже установлено. Следующее далее рассуждение соответствует второму из наших доказательств эргодической теоремы.

[2]) Здесь можно обойтись без ссылки на теорему 11.1, если, не вводя множество N_λ, заметить, что $\int\limits_{\{|x_n| > \lambda\}} (|x| + |y|)\,dP \underset{\lambda\to\infty}{\to} 0$ равномерно по n. — *Прим. ред.*

(сходимость по вероятности). Этот сильно ослабленный вариант равенства (11.4) можно доказать непосредственно с помощью энтропии. Если $\mathcal{A}$ — конечное поле с атомами A и A^c, то, как показывает теорема 7.1, $H(\mathcal{A}\,|\,\mathcal{A}_n)\to 0$. Но $H(\mathcal{A}\,|\,\mathcal{A}_n)$ есть математическое ожидание величины

$$\eta(P\{A\,\|\,\mathcal{A}_n\})+\eta(P\{A^c\,\|\,\mathcal{A}_n\}),$$

которая сходится, таким образом, по вероятности к нулю. Так как функция $\eta(t)$ равна нулю только в точках 0 и 1, то

$$P\{\varepsilon\leqslant P\{A\,\|\,\mathcal{A}_n\}\leqslant 1-\varepsilon\}\to 0.$$

Но

$$P(A\cap\{P\{A\,\|\mathcal{A}_n\}<\varepsilon\})=\int\limits_{\{P\{A\,\|\,\mathcal{A}_n\}<\varepsilon\}} P\{A\,\|\,\mathcal{A}_n\}\,dP\leqslant\varepsilon.$$

Из этих двух соотношений вытекает, что

$$P(A\cap\{P\{A\,\|\,\mathcal{A}_n\}\leqslant 1-\varepsilon\})<2\varepsilon$$

для больших n. Аналогично

$$P(A^c\cap\{P\{A^c\,\|\,\mathcal{A}_n\}\leqslant 1-\varepsilon\})<2\varepsilon$$

для больших n. Теперь легко получаем соотношение (11.10).

Пример 11.2. Пусть P — мера Лебега на классе $\mathcal{F}$ борелевских множеств в $\Omega=[0,\,1)$ и $\mathcal{G}_n$ — конечные поля, имеющие атомами диадические интервалы $[(k-1)/2^n,\ k/2^n)$ ранга n. Тогда $\mathcal{G}_n\uparrow\mathcal{F}$, так что, если $x(\omega)$ измерима по Борелю и интегрируема по Лебегу, то $E\{x\,\|\,\mathcal{G}_n\}\to x$ п. в. Вариантом величины $E\{x\,\|\,\mathcal{G}_n\}_\omega$ является $\left(\int\limits_{u_n(\omega)} x(\omega')\,d\omega'\right)\Big/2^{-n}$, где $u_n(\omega)$ — диадический интервал ранга n, содержащий ω. Таким образом,

$$\frac{1}{2^{-n}}\int\limits_{u_n(\omega)} x(\omega')\,d\omega'\to x(\omega)$$

для почти всех ω. По существу это основная теорема исчисления.

Пример 11.3. Пусть $\{x_n\}$ — координатные величины, соответствующие сдвигу с пространством состояний ρ. Если $\mathcal{G}_n$ есть σ-поле, порожденное величинами $x_{-n},\ \ldots,\ x_{-1}$, и $\mathcal{G}$ — поле, порожденное $\ldots,\ x_{-2},\ x_{-1}$, то $\mathcal{G}_n\uparrow\mathcal{G}$, так что в силу теоремы 11.2

$$\lim_{n\to\infty} P\{x_0=i\,\|\,x_{-n},\ \ldots,\ x_{-1}\}=P\{x_0=i\,\|\,\ldots,\ x_{-2},\ x_{-1}\}\quad\text{п. в.}$$

Если $\{x_n\}$ — марковская цепь, то это равенство очевидно в силу (9.8) и (9.9).

Пример 11.4. Теорема 11.2 вновь приводит нас к критерию эргодичности для марковского сдвига. Покажем сначала, что если множество A инвариантно, то $P(A+M)=0$ для некоторого M, лежащего во всех σ-полях $\mathcal{F}_n$, где $\mathcal{F}_n$ порождено $x_n, x_{n+1}, \ldots$. В самом деле, A может быть аппроксимировано (в смысле малости симметрической разности) цилиндром E. Но $T^{-m}E$ аппроксимирует $T^{-m}A=A$ и $T^{-m}E$ принадлежит $\mathcal{F}_n$, если m достаточно велико. Таким образом, A может быть аппроксимировано элементами из $\mathcal{F}_n$. Отсюда легко следует, что $P(A+B_n)=0$ для некоторого B_n из $\mathcal{F}_n$. Возьмем M равным, скажем, $\overline{\lim\limits_n} B_n$.

Так как M лежит в $\mathcal{F}_{n+1}$, то из очевидного обобщения равенства (9.13) вытекает, что

$$P\{M\|\ldots, x_{n-1}, x_n\}=P\{M\|x_n\} \quad \text{п. в.}$$

Так как $P(A+M)=0$, то A и M имеют равные почти всюду условные вероятности и, следовательно,

$$P\{A\|\ldots, x_{n-1}, x_n\}=P\{A\|x_n\} \quad \text{п. в.}$$

Так как σ-поля $\mathcal{G}_n$, порожденные последовательностями $\ldots, x_{n-1}, x_n$, удовлетворяют соотношению $\mathcal{G}_n\uparrow\mathcal{F}$, где $\mathcal{F}$ — полное σ-поле, порожденное всеми координатными величинами, то теорема 11.2 дает теперь

$$\lim_{n\to\infty} P\{A\|x_n\}=I_A \quad \text{п. в.}$$

Таким образом,

$$\lim_{n\to\infty} P\{|P\{A\|x_n\}-I_A|>\varepsilon\}=0. \tag{11.11}$$

В силу (10.6) и того, что A инвариантно, имеем

$$P\{A\|x_{n+1}\}_\omega=P\{A\|x_n\}_{T\omega}.$$

И, опять-таки в силу инвариантности A, имеем $I_A(\omega)=I_A(T\omega)$. Следовательно, если H_n есть множество, вероятность которого фигурирует в (11.11), то $T^{-1}H_n=H_{n+1}$. Так как T сохраняет P, то эта вероятность равна 0 для всех n. Таким образом,

$$P\{A\|x_0\}=I_A \quad \text{п. в.}$$

Следовательно, существует такой цилиндр B, зависящий только от координаты x_0, что $P(A+B)=0$. Таким образом,

T эргодично тогда и только тогда, когда не существует такого цилиндра B, что $P(B + T^{-1}B) = 0$ и $0 < P(B) < 1$. Это сразу приводит к критерию эргодичности из § 3: T эргодично тогда и только тогда, когда матрица вероятностей перехода Π неприводима (в предположении, что все стационарные вероятности p_i положительны). Этот вывод обладает тем преимуществом, что он проходит даже в случае, когда пространство состояний ρ не дискретно (см. Дуб [1]).

Убывающие σ-поля*

Если $\mathscr{G}_1 \subset \mathscr{G}_2 \subset \ldots$ и $x_n = E\{x \| \mathscr{G}_n\}$, то

$$E\{x_{n+1} \| x_1, \ldots, x_n\} = x_n \quad \text{п. в.}$$

Любая последовательность интегрируемых случайных величин, удовлетворяющих этому соотношению, называется *мартингалом*. Дубу принадлежит известная теорема о сходимости мартингалов почти всюду, которая содержит теорему 11.2 в качестве частного случая.

Предположим теперь, что $\mathscr{G}_1 \supset \mathscr{G}_2 \supset \ldots$ и $\mathscr{G} = \bigcap\limits_{n=1}^{\infty} \mathscr{G}_n$ (обозначаем это записью $\mathscr{G}_n \downarrow \mathscr{G}$). Следующий результат также вытекает из теоремы Дуба.

Если $\mathscr{G}_n \downarrow \mathscr{G}$, то

$$\lim_{n \to \infty} E\{x \| \mathscr{G}_n\} = E\{x \| \mathscr{G}\} \quad \text{п. в.} \tag{11.12}$$

для любой интегрируемой функции x и

$$\lim_{n \to \infty} P\{M \| \mathscr{G}_n\} = P\{M \| \mathscr{G}\} \quad \text{п. в.} \tag{11.13}$$

Мы не будем доказывать этот результат, но проиллюстрируем его, используя для доказательства перемешивающего свойства любого колмогоровского сдвига.

Напомним (см. § 8), что T является сдвигом Колмогорова, если каждое множество из σ-поля $\mathscr{F}_\infty = \bigcap\limits_{n=1}^{\infty} \mathscr{F}_n$ (где $\mathscr{F}_n$ порождено координатными величинами $x_n, x_{n+1}, \ldots$) имеет меру либо 0, либо 1. Для того чтобы доказать, что преобразование T перемешивающее, достаточно в силу теоремы 1.2 показать, что

$$\lim_{n \to \infty} P(A \cap T^{-n}B) = P(A)P(B), \tag{11.14}$$

если только B — цилиндр. Не ограничивая общности, будем предполагать, что B зависит только от неотрицательных координат, а в этом случае $T^{-n}B \in \mathscr{F}_n$ для всех n. Так как $\mathscr{F}_n \downarrow \mathscr{F}_\infty$, то в силу (11.13)

$$\lim_{n\to\infty} P\{A \| \mathscr{F}_n\} = P\{A \| \mathscr{F}_\infty\} \quad \text{п. в.}$$

Так как любое множество из $\mathscr{F}_\infty$ имеет меру либо 0, либо 1, то $P\{A\|\mathscr{F}_\infty\} = P(A)$ п. в., так что

$$\Delta_n = P\{A\|\mathscr{F}_n\} - P(A) \to 0 \quad \text{п. в.}$$

Но, так как $T^{-n}B \in \mathscr{F}_n$, то

$$\left| P\left(A \cap T^{-n}B\right) - P(A)P(B) \right| = \left| \int_{T^{-n}B} \Delta_n \, dP \right| \leqslant E\{|\Delta_n|\}.$$

Так как $|\Delta_n| \leqslant 2$, то $E\{|\Delta_n|\} \to 0$, что доказывает (11.14).

В конце § 4, посвященного непрерывным дробям, было показано, что если $\mathscr{G}_n$ есть σ-поле, порожденное $a_n(\omega)$, $a_{n+1}(\omega), \ldots$, то $\mathscr{G}_\infty = \bigcap_{n=1}^{\infty} \mathscr{G}_n$ содержит лишь множества меры 0 или 1. Только что приведенное рассуждение доказывает также, очевидно, и то, что преобразование, рассматриваемое в § 4, является перемешивающим.

Доказательство того, что $\mathscr{G}_\infty$ тривиально, содержащееся в § 4, опирается только на соотношение (4.15). Это легко приводит к следующему критерию. Предположим, что T — сдвиг, обладающий тем свойством, что существует такая положительная константа C, что

$$\frac{1}{C} P(B) \leqslant P(B|A) \leqslant CP(B) \tag{11.15}$$

для цилиндров

$$\begin{aligned} A &= \{x_m = i_m, \ \ldots, \ x_n = i_n\}, \\ B &= \{x_u = j_u, \ \ldots, \ x_v = j_v\}, \end{aligned} \tag{11.16}$$

где $m \leqslant n < u \leqslant v$. Тогда преобразование T является сдвигом Колмогорова и, следовательно, обладает свойством перемешивания[1]).

[1]) Тривиальность σ-поля $\mathscr{G}_\infty$ вытекает уже из одного левого неравенства (11.15). Если же выполнено только правое неравенство, то найдется такое конечное поле $\mathscr{A}$, что любое множество из $\mathscr{G}_\infty$ отличается от некоторого множества из $\mathscr{A}$ на множество меры 0. Так обстоит, например, дело в случае марковского сдвига. — *Прим. ред.*

Если T — сдвиг Маркова, соответствующий матрице вероятностей перехода Π с положительными элементами, то (11.15), разумеется, выполнено. Если матрица Π неприводима и апериодична, то существует такое натуральное число k, что Π^k имеет только положительные элементы. В этом случае (11.15) выполнено, если в (11.16) $u-n \geqslant k$, и нетрудно видеть, что наше доказательство по-прежнему проходит. Таким образом, мы получаем еще одно доказательство того, что сдвиг Маркова с неприводимой апериодической матрицей вероятностей перехода является сдвигом Колмогорова и, следовательно, перемешивающим.

Замечание. Соотношение [11.4] принадлежит Леви [2]. Теорию мартингалов см. в книге Дуба [1].

ГЛАВА 4

Сходимость энтропии

12. ОБОБЩЕНИЕ УСЛОВНОЙ ЭНТРОПИИ [1])

Определение

Если $\mathcal{A}$ — конечное подполе поля $\mathcal{F}$, а $\mathcal{G}$ — его σ-подполе, то положим

$$H(\mathcal{A}|\mathcal{G}) = E\left\{\sum_A \eta(P\{A\|\mathcal{G}\})\right\}, \tag{12.1}$$

где сумма берется по всем атомам поля $\mathcal{A}$, а η определяется, как обычно, соотношением (5.11). В случае когда $\mathcal{G}$ конечно, величина (12.1), которую мы будем называть условной энтропией поля $\mathcal{A}$ при заданном $\mathcal{G}$, сводится к выражению (6.1), использованному в гл. 2.

Тихе выбирает точку ω в соответствии с распределением P и сообщает экспериментатору, какие элементы поля $\mathcal{G}$ содержат ω, а какие нет. В соответствии с эвристическими рассуждениями предыдущей главы $P\{A\|\mathcal{G}\}$ есть новая вероятность множества A, так что энтропия $\sum_A \eta(P\{A\|\mathcal{G}\})$ является мерой остающейся неуверенности экспериментатора относительно исхода $\mathcal{A}$. Величина $H(\mathcal{A}|\mathcal{G})$ дает среднее значение этой остающейся неуверенности. Можно рассматривать $\mathcal{G}$ как сложный эксперимент. Если экспериментатор знает исход $\mathcal{G}$, то он все же до некоторой степени неуверен в исходе $\mathcal{A}$ и получает некоторую дополнительную информацию, когда узнает исход $\mathcal{A}$. Остающаяся неуверенность и эта дополнительная информация измеряются одной и той же величиной $H(\mathcal{A}|\mathcal{G})$.

Эта величина имеет другую форму:

$$H(\mathcal{A}|\mathcal{G}) = E\left\{-\sum_A I_A \ln P\{A\|\mathcal{G}\}\right\}. \tag{12.2}$$

Правая часть этого равенства в силу теоремы 10.1 равна (вспомним, что $E\{E\{x\|\mathcal{G}\}\} = E\{x\}$)

$$E\left\{E\left\{-\sum_A I_A \ln P\{A\|\mathcal{G}\}\|\mathcal{G}\right\}\right\} = \\ = E\left\{-\sum_A E\{I_A\|\mathcal{G}\} \ln P\{A\|\mathcal{G}\}\right\} = H(\mathcal{A}|\mathcal{G}).$$

---

[1]) Читатель может опустить этот параграф, если согласится предполагать некоторые сдвиги в последующих параграфах марковскими.

Свойства функции $H(\mathcal{A}|\mathcal{G})$

Это расширенное понятие условной энтропии обладает теми же свойствами, что и ранее рассмотренное:

(C_1) $\quad H(\mathcal{A}\vee\mathcal{B}|\mathcal{G})=H(\mathcal{A}|\mathcal{G})+H(\mathcal{B}|\mathcal{A}\vee\mathcal{G})$.

(C_2) $\quad H(\mathcal{A}|\mathcal{G})\leqslant H(\mathcal{B}|\mathcal{G})$, если $\mathcal{A}\subset\mathcal{B}$.

(C_3) $\quad H(\mathcal{A}|\mathcal{G}_1)\leqslant H(\mathcal{A}|\mathcal{G}_2)$, если $\mathcal{G}_1\supset\mathcal{G}_2$.

(C_4) $\quad H(\mathcal{A}\vee\mathcal{B}|\mathcal{G})\leqslant H(\mathcal{A}|\mathcal{G})+H(\mathcal{B}|\mathcal{G})$.

(C_5) $\quad H\left(T^{-1}\mathcal{A}\,|\,T^{-1}\mathcal{G}\right)=H(\mathcal{A}|\mathcal{G})$.

Докажем (C_1), (C_3) и (C_5), остальное получается, как раньше.

Согласно (10.11), имеем

$$P\{B\|\mathcal{A}\vee\mathcal{G}\}=\sum_A I_A\frac{P\{B\cap A\|\mathcal{G}\}}{P\{A\|\mathcal{G}\}}\quad \text{п. в.,} \tag{12.3}$$

где суммирование распространяется на атомы поля $\mathcal{A}$. Для любой точки ω все слагаемые в (12.3), кроме одного, обращаются в нуль. Поэтому

$$\ln P\{B\|\mathcal{A}\vee\mathcal{G}\}=\sum_A I_A\ln\frac{P\{A\cap B\|\mathcal{G}\}}{P\{A\|\mathcal{G}\}}\quad \text{п. в.} \tag{12.4}$$

Но

$$\sum_{A,B} I_{A\cap B}\ln P\{A\cap B\|\mathcal{G}\}=$$

$$=\sum_{A,B} I_AI_B\ln P\{A\|\mathcal{G}\}+\sum_{A,B} I_AI_B\ln\frac{P\{A\cap B\|\mathcal{G}\}}{P\{A\|\mathcal{G}\}}=$$

$$=\sum_A I_A\ln P\{A\|\mathcal{G}\}+\sum_B I_B\ln P\{B\|\mathcal{A}\vee\mathcal{G}\}.$$

Взяв математическое ожидание и применив (12.2), получим (C_1).

Для доказательства (C_3) используем неравенство Йенсена (10.10). Пусть $x=P\{A\|\mathcal{G}_1\}$. Так как функция η выпукла, то

$$E\{\eta(x)\|\mathcal{G}_2\}\leqslant\eta(E\{x\|\mathcal{G}_2\})=\eta(P\{A\|\mathcal{G}_2\}),$$

где равенство вытекает из теоремы 10.2. Переходя к математическим ожиданиям в обеих частях неравенства, получаем

$$E\{\eta(P\{A\|\mathcal{G}_1\})\}\leqslant E\{\eta(P\{A\|\mathcal{G}_2\})\}.$$

Суммирование по всем атомам поля $\mathcal{A}$ дает (C_3). Это рассуждение в скрытой форме повторяет доказательство свойства (A_3) из § 6.

Из (10.6) вытекает

$$\eta\left(P\left\{T^{-1}A \| T^{-1}\mathcal{G}\right\}_\omega\right) = \eta\left(P\left\{A \| \mathcal{G}\right\}_{T\omega}\right).$$

Интегрируя, получаем

$$E\left\{\eta\left(P\left\{T^{-1}A \| T^{-1}\mathcal{G}\right\}\right)\right\} = E\left\{\eta\left(P\left\{A \| \mathcal{G}\right\}\right)\right\},$$

а суммирование по атомам поля $\mathcal{A}$ дает (C_5).

Рассмотрим σ-поля $\mathcal{G}_n$ и $\mathcal{G}$.

Теорема 12.1. *Если* $\mathcal{G}_n \uparrow \mathcal{G}$, *то*

$$\lim_n H(\mathcal{A} | \mathcal{G}_n) = H(\mathcal{A} | \mathcal{G}).$$

Доказательство. В силу теоремы 11.2

$$\lim_{n\to\infty} P\{A \| \mathcal{G}_n\} = P\{A \| \mathcal{G}\} \quad \text{п. в.}$$

Так как функция $\eta(t)$ непрерывна и ограничена, то утверждение теоремы следует из определения (12.1).

Можно использовать теорему 12.1 для нового доказательства теоремы 7.1, представляющей основополагающий результат Колмогорова. Центральный этап этого доказательства состоит в доказательстве того, что если конечные поля $\mathcal{A}_n$ удовлетворяют требованию $\mathcal{A}_1 \subset \mathcal{A}_2 \subset \ldots$ и $\mathcal{A} \subset \bigvee_{n=1}^{\infty} \mathcal{A}_n$, то $\lim_n H(\mathcal{A} | \mathcal{A}_n) = 0$. В силу теоремы 12.1 и свойства (C_3)

$$\lim_{n\to\infty} H(\mathcal{A} | \mathcal{A}_n) = H\left(\mathcal{A} \,\middle|\, \bigvee_{n=1}^{\infty} \mathcal{A}_n\right) \leqslant H(\mathcal{A} | \mathcal{A}) = 0.$$

Сравните это с примером 11.1.

Из теоремы 12.1 и свойства (B_1) из § 6 следует

$$h(\mathcal{A}, T) = H\left(\mathcal{A} \,\middle|\, \bigvee_{i=1}^{\infty} T^{-i}\mathcal{A}\right). \tag{12.5}$$

Аналогично если T обратимо, то

$$h(\mathcal{A}, T) = H\left(\mathcal{A} \,\middle|\, \bigvee_{i=1}^{\infty} T^{i}\mathcal{A}\right). \tag{12.6}$$

В § 6 мы интерпретировали $h(\mathcal{A}, T)$ как приближенное количество информации, извлекаемой при проведении экспе-

римента $\mathcal{A}$, если известны результаты большой, но конечной серии предыдущих проведений этого эксперимента. *Теперь можно интерпретировать эту величину как точное значение количества информации, полученное от проведения эксперимента, при условии, что известна вся предыстория.*

Две специальные формулы[1])

Введем обозначения $\mathcal{A}^- = \bigvee_{i=1}^{\infty} T^{-i}\mathcal{A}$ и $\mathcal{B}^- = \bigvee_{i=1}^{\infty} T^{-i}\mathcal{B}$.

Теорема 12.2. *Пусть преобразование T обратимо. Если либо $\mathcal{B} \subset \mathcal{A}$, либо $\mathcal{A} \subset \mathcal{B}$, то*

$$\lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{k=0}^{n-1} T^k\mathcal{A} \mid \mathcal{B}^-\right) = H(\mathcal{A} \mid \mathcal{A}^-). \tag{12.7}$$

Доказательство. Рассмотрим сначала случай $\mathcal{B} \subset \mathcal{A}$. В силу свойства (C_1) (обобщенного по индукции) и (C_5)

$$\frac{1}{n} H\left(\bigvee_{k=0}^{n-1} T^k\mathcal{A} \mid \mathcal{B}^-\right) = \frac{1}{n}\sum_{k=0}^{n-1} H\left(T^k\mathcal{A} \mid \mathcal{B}^- \vee \bigvee_{i=0}^{k-1} T^i\mathcal{A}\right) =$$

$$= \frac{1}{n}\sum_{k=0}^{n-1} H\left(\mathcal{A} \mid T^{-k}\left(\mathcal{B}^- \vee \bigvee_{i=0}^{k-1} T^i\mathcal{A}\right)\right) =$$

$$= \frac{1}{n}\sum_{k=0}^{n-1} H(\mathcal{A} \mid T^{-1}\mathcal{A} \vee \ldots \vee T^{-k}\mathcal{A} \vee T^{-(k+1)}\mathcal{B} \vee T^{-(k+2)}\mathcal{B} \vee \ldots).$$

Если $\mathcal{B} \subset \mathcal{A}$, то теорема 12.1 дает

$$\lim_{k\to\infty} H(\mathcal{A} \mid T^{-1}\mathcal{A} \vee \ldots \vee T^{-k}\mathcal{A} \vee T^{-(k+1)}\mathcal{B} \vee \ldots) = H(\mathcal{A} \mid \mathcal{A}^-).$$

Теперь (12.7) следует из теоремы о средних арифметических.

Рассмотрим случай $\mathcal{A} \subset \mathcal{B}$. Так как $\mathcal{A}^- \subset \mathcal{B}^-$, то в силу (C_3) и первого случая (при $\mathcal{A} = \mathcal{B}$) имеем

$$\lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{k=0}^{n-1} T^k\mathcal{A} \mid \mathcal{B}^-\right) \leqslant \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{k=0}^{n-1} T^k\mathcal{A} \mid \mathcal{A}^-\right) = H(\mathcal{A} \mid \mathcal{A}^-). \tag{12.8}$$

[1]) Последние два результата этого параграфа используются только при доказательстве теоремы 17.3. Возможно, лучше будет вернуться к ним еще раз, когда они понадобятся.

С другой стороны, используя поочередно (C_1), (C_3) и (C_1), каждый раз применяя первый случай, получаем

$$\lim_{n\to\infty}\frac{1}{n}H\left(\bigvee_{k=0}^{n-1}T^k\mathcal{A}\mid\mathcal{B}^-\right)=$$

$$=\lim_{n\to\infty}\left[\frac{1}{n}H\left(\bigvee_{k=0}^{n-1}T^k\mathcal{B}\mid\mathcal{B}^-\right)-\frac{1}{n}H\left(\bigvee_{k=0}^{n-1}T^k\mathcal{B}\,\middle|\,\bigvee_{k=0}^{n-1}T^k\mathcal{A}\vee\mathcal{B}^-\right)\right]\geqslant$$

$$\geqslant H(\mathcal{B}\mid\mathcal{B}^-)-\lim_{n\to\infty}\frac{1}{n}H\left(\bigvee_{k=0}^{n-1}T^k\mathcal{B}\,\middle|\,\bigvee_{k=0}^{n-1}T^k\mathcal{A}\vee\mathcal{A}^-\right)=$$

$$=\lim_{n\to\infty}\left[\frac{1}{n}H\left(\bigvee_{k=0}^{n-1}T^k\mathcal{B}\mid\mathcal{A}^-\right)-\frac{1}{n}H\left(\bigvee_{k=0}^{n-1}T^k\mathcal{B}\,\middle|\,\bigvee_{k=0}^{n-1}T^k\mathcal{A}\vee\mathcal{A}^-\right)\right]=$$

$$=\lim_{n\to\infty}\frac{1}{n}H\left(\bigvee_{k=0}^{n-1}T^k\mathcal{A}\mid\mathcal{A}^-\right)=H(\mathcal{A}\mid\mathcal{A}^-).$$

Это вместе с (12.8) дает (12.7).

Теорема 12.3. *Если преобразование T обратимо, то*

$$H(\mathcal{A}\vee\mathcal{B}\mid\mathcal{A}^-\vee\mathcal{B}^-)=H\left(\mathcal{A}\mid\mathcal{A}^-\vee\bigvee_{i=-\infty}^{\infty}T^i\mathcal{B}\right)+H(\mathcal{B}\mid\mathcal{B}^-).$$

Доказательство. Теорема 12.1 дает

$$\lim_{n\to\infty}H\left(\mathcal{A}\mid\mathcal{A}^-\vee\mathcal{B}^-\vee\bigvee_{k=0}^{n-1}T^k\mathcal{B}\right)=H\left(\mathcal{A}\mid\mathcal{A}^-\vee\bigvee_{i=-\infty}^{\infty}T^i\mathcal{B}\right).\quad(12.9)$$

В силу (C_1) и (C_5)

$$H\left(\bigvee_{k=0}^{n-1}T^k\mathcal{A}\mid\mathcal{A}^-\vee\mathcal{B}^-\vee\bigvee_{i=0}^{n-1}T^i\mathcal{B}\right)=$$

$$=\sum_{k=0}^{n-1}H\left(T^k\mathcal{A}\mid\mathcal{A}^-\vee\mathcal{B}^-\vee\bigvee_{i=0}^{n-1}T^i\mathcal{B}\vee\bigvee_{j=0}^{k-1}T^j\mathcal{A}\right)=$$

$$=\sum_{k=0}^{n-1}H\left(\mathcal{A}\mid\mathcal{A}^-\vee\mathcal{B}^-\vee\bigvee_{i=0}^{n-k-1}T^i\mathcal{B}\right)=$$

$$=\sum_{k=0}^{n}H\left(\mathcal{A}\mid\mathcal{A}^-\vee\mathcal{B}^-\vee\bigvee_{i=0}^{k-1}T^i\mathcal{B}\right).$$

Отсюда, из (12.9) и теоремы о средних арифметических следует

$$\lim_{n\to\infty}\frac{1}{n}H\left(\bigvee_{k=0}^{n-1}T^k\mathcal{A}\mid\mathcal{A}^-\vee\mathcal{B}^-\vee\bigvee_{k=0}^{n-1}T^k\mathcal{B}\right)=H\left(\mathcal{A}\mid\mathcal{A}^-\vee\bigvee_{i=-\infty}^{\infty}T^i\mathcal{B}\right).$$

С другой стороны, в силу (C_1) и двукратного применения предыдущей теоремы (сначала с $\mathcal{A} \vee \mathcal{B}$ вместо $\mathcal{A}$ и $\mathcal{B}$, а затем с $\mathcal{B}$ вместо $\mathcal{A}$ и $\mathcal{A} \vee \mathcal{B}$ вместо $\mathcal{B}$) имеем

$$\lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{k=0}^{n-1} T^k\mathcal{A} \mid \mathcal{A}^- \vee \mathcal{B}^- \vee \bigvee_{k=0}^{n-1} T^k\mathcal{B}\right) =$$
$$= \lim_{n\to\infty}\left[\frac{1}{n} H\left(\bigvee_{k=0}^{n-1} T^k\mathcal{A} \vee \bigvee_{k=0}^{n-1} T^k\mathcal{B} \mid \mathcal{A}^- \vee \mathcal{B}^-\right) - \right.$$
$$\left. - \frac{1}{n} H\left(\bigvee_{k=0}^{n-1} T^k\mathcal{B} \mid \mathcal{A}^- \vee \mathcal{B}^-\right)\right] =$$
$$= H(\mathcal{A} \vee \mathcal{B} \mid \mathcal{A}^- \vee \mathcal{B}^-) - H(\mathcal{B} \mid \mathcal{B}^-).$$

З а м е ч а н и е. Теоремы 12.2 и 12.3 принадлежат Рохлину и Синаю [1].

13. ТЕОРЕМА ШЕННОНА — МАКМИЛЛАНА — БРЕЙМАНА

Если T — общий сдвиг (пример 1.2) с пространством состояний ρ, а $H(x_0, \ldots, x_{n-1})$ обозначает энтропию конечного поля, имеющего своими атомами r^n множеств (r есть объем множества ρ) вида $\{\omega : x_0(\omega) = i_0, \ldots, x_{n-1}(\omega) = i_{n-1}\}$, то в силу теоремы Колмогорова

$$h(T) = \lim_{n\to\infty} \frac{1}{n} H(x_0, \ldots, x_{n-1}). \tag{13.1}$$

Напомним, что (13.1) — теорема, а не определение.

Для любой последовательности $(i_0, \ldots, i_{n-1})$ элементов из ρ положим $p(i_0, \ldots, i_{n-1}) = P\{x_0 = i_0, \ldots, x_{n-1} = i_{n-1}\}$. Тогда $p(x_0(\omega), \ldots, x_{n-1}(\omega))$ есть вероятность фактически наблюдаемой последовательности $(x_0(\omega), \ldots, x_{n-1}(\omega))$. Так как $-\ln p(x_0(\omega), \ldots, x_{n-1}(\omega))$ имеет математическое ожидание, равное $H(x_0, \ldots, x_{n-1})$, то из (13.1) следует

$$\lim_{n\to\infty} E\left\{-\frac{1}{n} \ln p(x_0(\omega), \ldots, x_{n-1}(\omega))\right\} = h(T). \tag{13.2}$$

Результат

В эргодическом случае в соответствии со следующей теоремой, последовательно усиливаемые варианты которой принадлежат Шеннону, Макмиллану и Брейману, величина, стоящая под знаком математического ожидания в соотношении (13.2), сходится к $h(T)$ почти всюду. Для больших n вероятность $p(x_0(\omega), \ldots, x_{n-1}(\omega))$ наблюдаемой последовательности должна быть близка к $e^{-nh(T)}$.

Теорема 13.1. *Если T — эргодический сдвиг, то*

$$\lim_{n\to\infty}\left\{-\frac{1}{n}\ln p\,(x_0(\omega),\ \ldots,\ x_{n-1}(\omega))\right\}=h(T)\ \text{п. в.}\tag{13.3}$$

Доказательство. Разберем сначала два поучительных частных случая. Если T — сдвиг Бернулли $(p_1, \ldots, p_r)$, то $p(i_0, \ldots, i_{n-1}) = p_{i_0} \ldots p_{i_{n-1}}$, так что

$$-\frac{1}{n}\ln p\,(x_0(\omega),\ \ldots,\ x_{n-1}(\omega))=-\frac{1}{n}\sum_{k=0}^{n-1}\ln p_{x_k(\omega)}.$$

Из эргодической теоремы, которая в этом случае сводится к усиленному закону больших чисел для испытаний Бернулли, следует, что левая часть этого равенства сходится почти всюду к $E\{-\ln p_{x_0(\omega)}\}=-\sum_i p_i\ln p_i=h(T)$.

Если T — эргодический сдвиг Маркова, порожденнный начальными вероятностями p_i и вероятностями перехода p_{ij}, то

$$-\frac{1}{n}\ln p\,(x_0(\omega),\ \ldots,\ x_{n-1}(\omega))=$$
$$=-\frac{1}{n}\ln p_{x_0(\omega)}-\frac{1}{n}\sum_{k=1}^{n-1}\ln p_{x_{k-1}(\omega),\ x_k(\omega)}.$$

Соотношение (13.3) получается и в этом частном случае после применения эргодической теоремы.

Доказательство в общем случае опирается на эргодическую теорему и теорему 11.2 о сходимости условных вероятностей. Мы будем использовать функции

$$g_0(\omega)=-\ln p\,(x_0(\omega)),$$
$$g_k(\omega)=-\ln\frac{p\,(x_{-k}(\omega),\ \ldots,\ x_{-1}(\omega),\ x_0(\omega))}{p\,(x_{-k}(\omega),\ \ldots,\ x_{-1}(\omega))},$$
$$f_k^{(i)}(\omega)=-\ln\frac{p\,(x_{-k}(\omega),\ \ldots,\ x_{-1}(\omega),\ i)}{p\,(x_{-k}(\omega),\ \ldots,\ x_{-1}(\omega))}.$$

Заметим, что $f_k^{(i)}(\omega)$ есть взятый с обратным знаком логарифм величины

$$P\{x_0=i\,\|\,x_{-k},\ \ldots,\ x_{-1}\}_\omega.$$

Заметим также, что все эти функции неотрицательны.

Непосредственное вычисление показывает, что

$$-\frac{1}{n}\ln p\,(x_0(\omega),\ \ldots,\ x_{n-1}(\omega))=\frac{1}{n}\sum_{k=0}^{n-1}g_k(T^k\omega).\tag{13.4}$$

Если бы функция $g_k(\omega)$ не зависела от k, то правая часть равенства (13.4) имела бы вид среднего, к которому приме-

няется эргодическая теорема, и мы получили бы (13.3). Хотя, вообще говоря, $g_k(\omega)$ зависит от k, мы покажем, используя теорему (11.2), что $g_k(\omega)$ сходится к некоторой функции $g(\omega)$ при $k \to \infty$. После этого идея доказательства заключается в том, чтобы проверить, что среднее в (13.4) близко к $\frac{1}{n}\sum_{k=0}^{n-1} g(T^k\omega)$. (Бернуллиевский случай разбирается просто, поскольку для всех неотрицательных k функции $g_k(\omega)$ совпадают со своим пределом $g(\omega)$. Марковский случай прост, так как $g_k(\omega) = g(\omega)$ для положительных k.)

В силу теоремы 11.2 $P\{x_0 = i \| x_{-k}, \ldots, x_{-1}\}$ сходится почти всюду к

$$P\{x_0 = i \| \ldots, x_{-2}, x_{-1}\}.$$

Из непрерывности логарифма следует, что $f_k^{(i)}(\omega)$ сходится п. в. (предел может быть равен $+\infty$). Так как $g_k(\omega)$ совпадает с $f_k^{(i)}(\omega)$ на цилиндре $\{\omega : x_0(\omega) = i\}$, то предел

$$g(\omega) = \lim_{k \to \infty} g_k(\omega) \tag{13.5}$$

существует п. в. Так как функции $g_k(\omega)$ неотрицательны, то $g(\omega)$ также неотрицательна, но то, что мы пока знаем, не противоречит предположению о том, что на множестве положительной меры она может равняться $+\infty$.

Покажем теперь, что

$$E\left\{\sup_k g_k(\omega)\right\} < \infty. \tag{13.6}$$

В частности, получим, что $g(\omega)$ интегрируема и, следовательно, конечна п. в. Если

$$E_k = \left\{\omega : \max_{1 \leqslant j < k} g_j(\omega) \leqslant \lambda < g_k(\omega)\right\},$$

то

$$P(E_k) = \sum_i P(\{x_0 = i\} \cap E_k) = \sum_i P\left(\{x_0 = i\} \cap F_k^{(i)}\right),$$

где

$$F_k^{(i)} = \left\{\omega : \max_{i \leqslant j < k} f_j^{(i)}(\omega) \leqslant \lambda < f_k^{(i)}(\omega)\right\}.$$

Так как $F_k^{(i)}$ принадлежит σ-полю, порожденному величинами $x_{-k}, \ldots, x_{-1}$, то

$$P\left(\{x_0 = i\} \cap F_k^{(i)}\right) = \int_{F_k^{(i)}} P\{x_0 = i \| x_{-k} \ldots x_{-1}\}\, dP =$$

$$= \int_{F_k^{(i)}} e^{-f_k^{(i)}(\omega)} P(d\omega) \leqslant e^{-\lambda} P\left(F_k^{(i)}\right).$$

Так как $F_k^{(i)}$ при различных k не пересекаются, то

$$\sum_k P(E_k) \leqslant \sum_i e^{-\lambda} \sum_k P\left(E_k^{(i)}\right) \leqslant re^{-\lambda},$$

где r — объем множества ρ. Следовательно,

$$P\left\{\omega\colon \sup_k g_k(\omega) > \lambda\right\} \leqslant re^{-\lambda},$$

откуда следует (13.6). Таким образом, g интегрируема.

Интегрирование соотношения (13.4) и замена переменной дают

$$E\{g\} = \lim_{k\to\infty} E\{g_k\} = \lim_{n\to\infty} E\left\{\frac{1}{n}\sum_{k=0}^{n-1} g_k(T^k\omega)\right\} = h(T).$$

Запишем

$$\frac{1}{n}\sum_{k=0}^{n-1} g_k(T^k\omega) = \frac{1}{n}\sum_{k=0}^{n-1} g(T^k\omega) + \frac{1}{n}\sum_{k=0}^{n-1}(g_k(T^k\omega) - g(T^k\omega)). \quad (13.7)$$

По эргодической теореме (вспомним, что T эргодично)

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} g(T^k\omega) = E\{g\} = h(T) \text{ п. в.} \quad (13.8)$$

Если $G_N(\omega) = \sup_{k\geqslant N} |g_k(\omega) - g(\omega)|$, то для любого N

$$\varlimsup_{n\to\infty} \left|\frac{1}{n}\sum_{k=0}^{n-1}(g_k(T^k\omega) - g(T^k\omega))\right| \leqslant$$

$$\leqslant \varlimsup_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} |g_k(T^k\omega) - g(T^k\omega)| \leqslant$$

$$\leqslant \varlimsup_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} G_N(T^k\omega) = E\{G_N\} \text{ п. в.}$$

Но $G_N(\omega)$ сходится к нулю почти всюду и мажорируется интегрируемой функцией $g(\omega) + \sup_k g_k(\omega)$, так что $\lim_{N\to\infty} E\{G_N\} = 0$. Значит, вторая сумма в правой части равенства (13.7) стремится к нулю почти всюду, что вместе с (13.8) дает

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} g_k(T^k\omega) = h(T) \text{ п. в.}$$

Теперь (13.3) следует из (13.4), что завершает доказательство теоремы.

Другие варианты теоремы

Если T не обязательно эргодично, то (13.8) нужно заменить соотношением

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} g(T^k \omega) = \hat{g}(\omega) \text{ п. в.},$$

где $\hat{g}$ — инвариантная функция со средним значением $E\{\hat{g}\} = h(T)$. Аналогично следующее за (13.8) неравенство должно быть заменено неравенством

$$\varlimsup_{n \to \infty} \left| \frac{1}{n} \sum_{k=0}^{n-1} (g_k(T^k\omega) - g(T^k\omega)) \right| \leqslant \widehat{G}_N(\omega) \text{ п. в.},$$

где теперь $G_N(\omega) = \sup\limits_{k \geqslant n} |g_k(\omega) - g(\omega)|$ и $E\{\widehat{G_N}\} = E\{G_N\}$. Левая часть этого неравенства превосходит λ с вероятностью, не превосходящей $E\{G_n\}/\lambda$. Полагая $N \to \infty$, $\lambda \to 0$, получаем, что она равна 0 почти всюду. Отсюда следует, что

$$\lim_{n \to \infty} \left\{ -\frac{1}{n} \ln p(x_0(\omega), \ldots, x_{n-1}(\omega)) \right\} = \hat{g}(\omega) \text{ п. в.}$$

Если T эргодично, то мы снова приходим к (13.3).

Теорема 13.1 сохраняет силу и для одностороннего сдвига T. Если задан односторонний эргодический сдвиг T, то построим двусторонний сдвиг $\tilde{T}$ с теми же конечномерными мерами. Так как эргодичность связана только с конечномерными мерами, то $\tilde{T}$ также эргодичен. Отображение, переводящее бесконечную в обе стороны последовательность $(\ldots, \omega_{-1}, \omega_0, \omega_1, \ldots)$ в последовательность $(\omega_0, \omega_1, \ldots)$, переводит соотношение (13.3) для сдвига $\tilde{T}$ в соответствующее соотношение для T. Таким образом, теорема 13.1 выполнена и для односторонних сдвигов.

Из теоремы 13.1 для одностороннего сдвига можно вывести более общий на вид результат. Пусть T — произвольное эргодическое, сохраняющее меру преобразование на вероятностном пространстве $(\Omega, \mathcal{F}, P)$, а $\mathcal{A}$ — конечное подполе поля $\mathcal{F}$. Если $p_n(\omega)$ — вероятность того атома поля $\bigvee\limits_{k=0}^{n-1} T^{-k}\mathcal{A}$, который содержит ω, то

$$\lim_{n \to \infty} \left\{ -\frac{1}{n} \ln p_n(\omega) \right\} = h(\mathcal{A}, T) \text{ п. в.} \tag{13.9}$$

$\left(\text{Если } \bigvee_{k=0}^{\infty} T^{-k}\mathcal{A} = \mathcal{F}, \text{ то } h(\mathcal{A}, T) \text{ можно отождествить с } h(T).\right)$ Поэтому нужно только перенести задачу на пространство односторонних последовательностей с помощью отображения $\omega \to (f(\omega), f(T\omega), \ldots)$, где f — некоторая функция, принимающая различные значения на разных атомах поля $\mathcal{A}$. С помощью (13.9) можно, полагая $\mathcal{A}$ полем событий, наблюдаемых в момент времени 0, снова прийти к теореме 13.1.

Аналогичный результат имеет место, если $\mathcal{A}$ заменить σ-полем, соответствующим счетному (а не конечному) $\mathcal{F}$-разбиению пространства Ω. Если $p_n(\omega)$ — по-прежнему вероятность того атома поля $\bigvee_{k=0}^{n-1} T^{-k}\mathcal{A}$, который содержит ω, и $\bigvee_{k=0}^{\infty} T^{-k}\mathcal{A} = \mathcal{F}$, то

$$\lim_{n\to\infty}\left\{-\frac{1}{n}\ln p_n(\omega)\right\} = h(T) \text{ п. в.}^{1)}$$

Отсюда следует, например, что преобразование, связанное с непрерывными дробями, имеет энтропию $\pi^2/6\ln 2$ (см. (4.22)).

Теорема Шеннона — Макмиллана — Бреймана связана с условиями Липшица, которым удовлетворяют некоторые функции распределения на единичном интервале. Возьмем в качестве T преобразование $T\omega = r\omega \pmod 1$ на полуинтервале $[0, 1]$ (пример 3.5), и пусть P — мера, сохраняемая преобразованием T. Пусть $\mathcal{A}$ — конечное поле с атомами $[i/r, (i+1)/r)$, $i = 0, 1, \ldots, r-1$. В этом случае $p_n(\omega)$ есть P-мера того r-адического интервала $u_n(\omega)$ ранга n, который содержит ω. Если T эргодично относительно P, то из (13.9) и того, что $\bigvee_{k=0}^{\infty} T^{-k}\mathcal{A}$ есть σ-поле $\mathcal{F}$ всех борелевских множеств, вытекает, что

$$\lim_{n\to\infty}\frac{\ln p_n(\omega)}{\ln r^{-n}} = \frac{h(T)}{\ln r} \tag{13.10}$$

с точностью до множества P-меры 0. Если (13.10) выполняется в точке ω, то

$$\lim_{n\to\infty}\frac{p_n(\omega)}{(r^{-n})^{\alpha}} = \begin{cases} \infty, & \text{если } \alpha > \dfrac{h(T)}{\ln r}, \\ 0, & \text{если } \alpha < \dfrac{h(T)}{\ln r}. \end{cases} \tag{13.11}$$

---

1) Доказательство см. в работе Чжун Кай-лая [1].

Если $F(x)=P[0, x)$, то $p_n(\omega)$ равно приращению функции распределения F на интервале $u_n(\omega)$. Так как этот интервал имеет длину r^{-n}, то (13.11) означает, что F удовлетворяет условию Липшица в точности порядка $h(T)/\ln r$ в точке ω. Таким образом, F удовлетворяет условию Липшица в точности порядка $h(T)/\ln r$ на множестве P-меры 1. (В § 3 отмечалось, что F сингулярна, если только не имеет места тождество $F(x)\equiv x$, в последнем случае $h(T)/\ln r=1$.) Установить условие Липшица для F на множестве лебеговой меры 1 представляется затруднительным.

Свойство равнораспределенности

Пусть T — снова двусторонний сдвиг. Для любого натурального b отображение $\omega\to(x_1(\omega), \ldots, x_b(\omega))$ индуцирует вероятностную меру на множестве ρ^b упорядоченных групп из b элементов пространства ρ, причем вероятность такой упорядоченной группы $u=(i_1, \ldots, i_b)$ равна

$$p(u)=P\{(x_1, \ldots, x_b)=u\}=P\{(x_{n+1}, \ldots, x_{n+b})=u\}.$$

Согласно теореме 13.2, при больших b множество ρ^b разлагается на два подмножества, из которых одно имеет малую общую вероятность, а другое состоит из упорядоченных групп по b элементов, таких, что вероятность каждой группы близка к $e^{-bh(T)}$ — факт, известный под названием свойства асимптотической равнораспределенности.

Теорема 13.2. *Пусть T — эргодический сдвиг с энтропией h. Тогда для любого $\varepsilon>0$ существует такое натуральное число $b_0(\varepsilon)$, что ρ^b при $b\geqslant b_0(\varepsilon)$ разлагается на два множества H и L со свойствами*:

1) $\sum\limits_{u\in L} p(u)=P\{(x_1, \ldots, x_b)\in L\}<\varepsilon$,

2) $e^{-b(h+\varepsilon)}<p(u)=P\{(x_1, \ldots, x_b)=u\}<e^{-b(h-\varepsilon)}$, *если* $u\in H$.

Доказательство. Перейдя в (13.3 к пределу по вероятности, получим

$$p\lim_{b\to\infty}\left\{-\frac{1}{b}\ln p(x_1(\omega), \ldots, x_b(\omega))\right\}=h.$$

Выберем такое $b_0(\varepsilon)$, что

$$P\left\{\omega:\left|-\frac{1}{b}\ln p(x_1(\omega), \ldots, x_b(\omega))-h\right|\geqslant\varepsilon\right\}<\varepsilon$$

при $b \geqslant b_0(\varepsilon)$. Пусть H (группа высокой вероятности) состоит из тех упорядоченных групп u по b элементов, для которых

$$\left| -\frac{1}{b} \ln p(u) - h \right| < \varepsilon,$$

и пусть L (группа низкой вероятности) является дополнением множества H в ρ^b. Ясно, что H и L обладают требуемыми свойствами.

Замечание. Исходные работы в этой области принадлежат Шеннону [1], Макмиллану [1] и Брейману [1]. Другое доказательство теоремы 13.2 см. в работе Томасяна [1].

14. СВЯЗЬ С ТЕОРИЕЙ РАЗМЕРНОСТИ *

Классическое определение

В 1919 г. Хаусдорфом было введено понятие размерности, которое, как оказалось, имеет любопытную связь с энтропией.

Пусть M — множество в метрическом пространстве, скажем в евклидовом. *Внешняя мера* $l_\alpha(M)$ (α-мерная) определяется для положительного α следующим образом. Назовем ρ-покрытием множества M счетное покрытие этого множества замкнутыми шарами S_i диаметра, меньшего ρ. Положим

$$l_\alpha(M, \rho) = \inf \sum_i (\text{диам. } S_i)^\alpha, \tag{14.1}$$

где нижняя грань берется по всем ρ-покрытиям множества M. При убывании ρ нижняя грань в (14.1) распространяется на все меньший класс покрытий и, следовательно, $l_\alpha(M, \rho)$ возрастает или, во всяком случае, не убывает. Поэтому предел (конечный или бесконечный)

$$l_\alpha(M) = \lim_{\rho \to 0} l_\alpha(M, \rho)$$

существует.

Ясно, что функция $l_\alpha(\cdot)$ монотонна: $l_\alpha(M) \leqslant l_\alpha(M')$, если $M \subset M'$. Для заданной последовательности множеств M_n выбираем такие ρ-покрытия $\{S_{ni}\}$, что $\sum_i (\text{диам. } S_{ni})^\alpha < < l_\alpha(M_n, \rho) + \varepsilon/2^n \leqslant l_\alpha(M_n) + \varepsilon/2^n$. Все шары S_{ni} вместе образуют ρ-покрытие множества $\bigcup_n M_n$, причем $\sum_{in} (\text{диам. } S_{ni})^\alpha < < \sum_n l_\alpha(M_n) + \varepsilon$. Следовательно, функция $l_\alpha(\cdot)$ полуаддитивна:

$l_\alpha\left(\bigcup_n M_n\right) \leqslant \sum_n l_\alpha(M_n)$. Таким образом, $l_\alpha(M)$ как функция от M является внешней мерой.

Хаусдорфова размерность множества M определяется поведением $l_\alpha(M)$ не как функции от M, а как функции от α. Покажем, что если $l_\alpha(M)$ конечно, то $l_{\alpha'}(M)=0$ для любого $\alpha' > \alpha$. Если $\{S_i\}$ — некоторое ρ-покрытие множества M, для которого

$$\sum_i (\text{диам. } S_i)^\alpha \leqslant l_\alpha(M, \rho) + 1 \leqslant l_\alpha(M) + 1 = K < \infty,$$

то $l_{\alpha'}(M, \rho) \leqslant \sum_i (\text{диам. } S_i)^{\alpha'} \leqslant \rho^{\alpha'-\alpha} \sum_i (\text{диам. } S_i)^\alpha < \rho^{\alpha'-\alpha} K$. Так как $\alpha' > \alpha$, то, полагая $\rho \to 0$, получаем $l_{\alpha'}(M)=0$. Если $l_\alpha(M)$ конечна для некоторого конкретного значения α, то

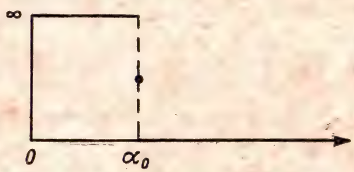


Рис. 4.

она равна нулю для всех бо́льших значений. Следовательно, существует „точка перехода“ — такая точка α_0, что $l_\alpha(M)=\infty$ для $\alpha < \alpha_0$ и $l_\alpha(M)=0$ для $\alpha > \alpha_0$ (рис. 4). Функция $l_\alpha(M)$ в точке α_0 может равняться нулю, принимать конечное положительное значение или ∞. (Крайние случаи $\alpha_0 = 0$ ($l_\alpha(M)=0$ для всех $\alpha > 0$) и $\alpha_0=\infty$ ($l_\alpha(M)=\infty$ для всех $\alpha>0$) могут иметь место, хотя в евклидовом пространстве второй случай невозможен.) Однозначно определяемое число α_0 и есть хаусдорфова размерность множества M, мы ее обозначаем $\dim M$.

Имеем

$$\dim M = \sup\{\alpha : l_\alpha(M)=\infty\} = \inf\{\alpha : l_\alpha(M)=0\}.$$

Мы постоянно используем следующие четыре факта: 1) $l_\alpha(M) > 0 \Rightarrow \dim M \geqslant \alpha$; 2) $\dim M \geqslant \alpha \Rightarrow l_\alpha(M) = \infty$; 3) $l_\alpha(M) < \infty \Rightarrow \dim M \leqslant \alpha$; 4) $\dim M < \alpha \Rightarrow l_\alpha(M)=0$.

Отметим два основных свойства хаусдорфовой размерности. Прежде всего, она, очевидно, монотонна:

$$\dim M \leqslant \dim M', \tag{14.2}$$

если $M \subset M'$. Во-вторых,

$$\dim \bigcup_n M_n = \sup_n \dim M_n. \tag{14.3}$$

Действительно, если $\dim M_n < \alpha$ для всех n, то $l_\alpha(M_n) = 0$ и в силу полуаддитивности функции $l_\alpha(\cdot)$ имеем $l_\alpha\left(\bigcup_n M_n\right) = 0$ и, следовательно, $\dim \bigcup_n M_n \leqslant \alpha$. Таким образом, $\dim \bigcup_n M_n \leqslant \sup_n \dim M_n$. Обратное неравенство очевидным образом следует из (14.2).

Для оправдания хаусдорфова определения покажем, что размерность достаточно гладкой поверхности в трехмерном пространстве равна 2. Итак, предположим, что поверхность M определяется уравнением $z = f(x, y)$ над единичным квадратом на плоскости. Без всяких дальнейших предположений можно доказать, что $\dim M \geqslant 2$. Для этого будем считать, что M покрыто шарами S_i диаметра d_i. Если p — вертикальная проекция на плоскость (x, y), то pS_i покрывают pM. Так как pS_i есть круг диаметра d_i (с площадью $\pi d_i^2/4$), а pM есть единичный квадрат (с площадью 1), то $\sum_i \pi d_i^2/4 \geqslant 1$, или $\sum_i (\text{диам. } S_i)^2 \geqslant 4/\pi$. Таким образом, $l_2(M) > 0$, так что $\dim M \geqslant 2$.

Обратное неравенство можно доказать, если предположить выполненным условие Липшица $\omega_f(\delta) = O(\delta)$, где

$$\omega_f(\delta) = \sup\{|f(x, y) - f(x', y')| : |x - x'| \leqslant \delta,\ |y - y'| \leqslant \delta\}$$

есть модуль непрерывности функции f. Из этого условия следует непрерывность f; с другой стороны, если, например, f непрерывна и имеет ограниченные частные производные, то это условие выполнено. Докажем, что $\dim M \leqslant 2$, показав, что $l_{2+\varepsilon}(M) = 0$ при любом $\varepsilon > 0$. Выберем $K > 1$ так, что $\omega_f(\delta) < K\delta$. Разобьем теперь единичный квадрат на n^2 маленьких квадратов со сторонами, равными $1/n$. Так как вариация функции f на любом из этих маленьких квадратов меньше, чем K/n, то часть поверхности M, лежащая над таким маленьким квадратом, может быть заключена в куб с ребром K/n, который в свою очередь может быть заключен в шар диаметра $\sqrt{3}K/n$. Таким образом, поверхность M может быть покрыта n^2 шарами диаметра $\sqrt{3}K/n$. Для этого

покрытия имеем $\sum_i (\text{диам. } S_i)^{2+\varepsilon} = (\sqrt{3}K)^{2+\varepsilon}/n^\varepsilon$. Выбирая n так, что $\sqrt{3}K/n < \rho$, получаем

$$l_{2+\varepsilon}(M, \rho) \leqslant (\sqrt{3}\,K)^{2+\varepsilon}/n^\varepsilon$$

и, полагая $n \to \infty$, имеем $l_{2+\varepsilon}(M, \rho) = 0$. Таким образом, $l_{2+\varepsilon}(M) = 0$, так что $\dim M \leqslant 2$.

Такое рассуждение показывает, что хорошие множества имеют именно ту размерность, которую и следует ожидать. Далее мы не будем заниматься подобными рассмотрениями, так как нас будут интересовать крайне нерегулярные множества, размерность которых не есть целое число.

Размерность в единичном интервале

Далее множество M будет подмножеством единичного интервала. Нетрудно видеть, что для такого множества $l_1(M)$ есть обычная внешняя мера, так что $l_1(M) \leqslant 1$. Таким образом, $\dim M$ лежит между 0 и 1. Если M — борелевское множество положительной лебеговой меры, то $l_1(M) > 0$, так что $\dim M = 1$. С другой стороны, каждое одноточечное множество, а следовательно, и каждое счетное множество имеет нулевую размерность[1]). Между этими крайними случаями лежит канторово множество, которое, как показал Хаусдорф, имеет размерность $\ln 2/\ln 3$. (Мы докажем это ниже.) Хаусдорфова размерность измеряет величину множеств на единичном интервале таким образом, что дает возможность решить вопрос о том, какое из двух множеств меры 0 „больше".

Пусть $M(p)$ — множество точек на единичном интервале, содержащих 1 в своих диадических разложениях в пропорции p, т. е. $\omega = 0, \omega_1\omega_2 \ldots$ лежит в $M(p)$ в том и только том случае, когда $\lim_n n^{-1} \sum_{k=1}^n \omega_k = p$. Так как $M(1/2)$ есть множество чисел, нормальных по основанию 2, то его лебегова мера равна 1. Если $p \neq 1/2$, то $M(p)$ состоит из чисел, не являющихся нормальными и, значит, имеет меру 0. Поэтому интересно узнать его размерность[2]). В 1949 г. Эгглстон доказал, что

$$\dim M(p) = -\frac{1}{\ln 2}[p \ln p + (1-p)\ln(1-p)].$$

---

[1]) Так как интервал есть объединение одноточечных множеств, то соотношение (14.3) не имеет места для несчетных сумм.

[2]) В силу (14.3) и того, что множество рациональных чисел имеет размерность 0, вопрос о том, конечно или бесконечно разложение рационального ω, не влияет на эту размерность.

Этот результат тесно соприкасается с хаусдорфовым результатом относительно канторова множества.

Более общим образом, фиксируем основание r, и пусть $x_n(\omega)$ есть n-й знак в разложении ω по основанию r. Таким образом, $\omega = \sum_{n=1}^{\infty} x_n(\omega)/r^n$. Пусть число $N_i(\omega, n)$ указывает, сколько раз i встречается среди знаков $x_1(\omega), \ldots, x_n(\omega)$. Для вероятностного вектора $(p_0, \ldots, p_{r-1})$ пусть $M(p_0, \ldots, p_{r-1})$ обозначает множество тех ω, для которых $\lim_n N_i(\omega, n)/n = p_i$, $i = 0, 1, \ldots, r-1$. Эгглстон показал, что

$$\dim M(p_0, \ldots, p_{r-1}) = -\frac{1}{\ln r} \sum_{i=0}^{r-1} p_i \ln p_i. \tag{14.4}$$

Мы дадим доказательство, демонстрирующее связь размерности с энтропией.

Обобщенное определение

Для истолкования (14.4) дадим сначала другое (эквивалентное) определение хаусдорфовой размерности, а затем обобщим его. Для единичного интервала каждый шар есть некоторый интервал, а его диаметр — длина этого интервала. Поэтому (14.1) сводится к равенству

$$l_\alpha(M, \rho) = \inf \sum_i |v_i|^\alpha,$$

где нижняя грань берется по всем покрытиям множества M интервалами v_i длины $|v_i| < \rho$. Далее r-аддический интервал

$$v = \left[\frac{j}{r^n}, \frac{j+1}{r^n}\right), \quad n = 1, 2, \ldots, \quad j = 0, 1, \ldots, r^n - 1,$$

будем называть цилиндром, так как он имеет вид

$$\{\omega : x_k(\omega) = i_k, \quad k = 1, \ldots, n\}$$

при подходящих i_k. (Здесь и далее r есть некоторое фиксированное основание.) Если

$$\lambda_\alpha(M, \rho) = \inf \sum_i |v_i|^\alpha, \tag{14.5}$$

где нижняя грань берется теперь только по покрытиям множества M *цилиндрами* длины, меньшей ρ, то $\lambda_\alpha(M, \rho)$ отличается от $l_\alpha(M, \rho)$, но это отличие не сказывается при

вычислении размерностей. Действительно, мы покажем, что

$$l_\alpha(M, \rho) \leqslant \lambda_\alpha(M, \rho) \leqslant 2rl_\alpha(M, \rho), \tag{14.6}$$

откуда будет следовать, что если $\lambda_\alpha(M) = \lim\limits_{\rho\to 0} \lambda_\alpha(M, \rho)$, то $\lambda_\alpha(M) = \infty$ при $\alpha < \dim M$ и $\lambda_\alpha(M) = 0$ при $\alpha > \dim M$, так что можно определить $\dim M$ через $\lambda_\alpha(M)$ так же, как и через $l_\alpha(M)$.

Левое из неравенств (14.6) очевидно. Докажем правое неравенство для случая $r = 2$ (доказательство в общем случае проводится аналогично). Достаточно показать, что если u — произвольный интервал, то существуют четыре цилиндра, каждый длины, не превосходящей $|u|$, покрывающие u. Выберем внутри u цилиндр v_1 максимальной длины $|v_1| = (1/2)^n$, так что u уже не содержит цилиндров длины $(1/2)^{n-1}$. Если v_0 и v_2 — цилиндры длины $(1/2)^n$, лежащие слева и справа от v_1 соответственно, то один из интервалов $v_0 \cup v_1$ или $v_1 \cup v_2$ является цилиндром длины $(1/2)^{n-1}$. Для определенности положим, что это интервал $v_0 \cup v_1$. Так как он не может лежать в u, то v_0 заходит за левый конец интервала u. Если v_3 — цилиндр длины $(1/2)^n$, лежащий справа от v_2, то $v_2 \cup v_3$ есть цилиндр длины $(1/2)^{n-1}$ и, следовательно, он не может лежать в u. Таким образом, v_3 простирается за правый конец интервала u и $v_0, \ldots, v_3$ покрывают u. При этом $|v_i| = (1/2)^n \leqslant |u|$.

Итак, можно определить $\dim M$ как такое число α_0, что $\lambda_\alpha(M) = \infty$ для $\alpha < \alpha_0$ и $\lambda_\alpha(M) = 0$ для $\alpha > \alpha_0$. Это новое определение полезно для вопросов, связанных с r-адическими разложениями, поскольку в нем фигурируют только покрытия r-адическими интервалами (или цилиндрами, как мы их только что называли)[1]).

Придав новую форму определению, обобщим его. Пусть μ — вероятностная мера на борелевских множествах единичного интервала. Положим

$$\mu_\alpha(M, \rho) = \inf \sum_i \mu(v_i)^\alpha, \tag{14.7}$$

где нижняя грань берется по μ-ρ-покрытиям множества M, т. е. по покрытиям цилиндрами v_i с мерой $\mu(v_i) < \rho$. В даль-

[1]) Определим размерность исходя из формулы (14.5), только теперь нижнюю грань будем брать по ρ-покрытиям интервалами из заданного класса $\mathcal{I}$. При каких условиях, налагаемых на $\mathcal{I}$, определенная таким образом размерность совпадает с хаусдорфовой? Мы доказали это совпадение для случая, когда $\mathcal{I}$ — класс r-адических интервалов, но мне неизвестны никакие общие условия. Результат такого рода, связанный с непрерывными дробями, см. Кинни и Питчер [1].

нейшем мы будем для простоты предполагать, что мера μ неатомическая; в противном случае μ-ρ-покрытия множества M могло бы и не существовать вовсе. Если μ — мера Лебега, которую мы будем обозначать λ, то (14.7) сводится к (14.5). При $\rho \to 0$ функция $\mu_\alpha(M, \rho)$ монотонно стремится к пределу $\mu_\alpha(M)$. Как и раньше, можно показать, что $\mu_\alpha(M)$ как функция от M есть внешняя мера и что для фиксированного M существует такое α_0, что $\mu_\alpha(M) = \infty$ при $\alpha < \alpha_0$ и $\mu_\alpha(M) = 0$ при $\alpha > \alpha_0$. Это α_0 называют размерностью множества M относительно μ и обозначают $\dim_\mu M$.

Ясно, что $\dim_\mu M$ совпадает с исходной $\dim M$ для любого M из единичного интервала. Для любого μ имеем $0 \leqslant \dim_\mu M \leqslant 1$; кроме того, $\dim_\mu M = 0$ для счетных M и $\dim_\mu M = 1$ для борелевских множеств M, для которых $\mu(M) > 0$. Соотношения, аналогичные (14.2) и (14.3), также выполняются.

Основной результат

Пусть $u_n(\omega)$ — цилиндр длины $(1/r)^n$, содержащий ω, т. е. тот единственный интервал вида $\{\omega : x_k(\omega) = i_k,\ k = 1, \ldots, n\}$, который содержит ω. Заметим, что ω' лежит в $u_n(\omega)$ в том и только том случае, когда разложения ω и ω' по основанию r совпадают на первых n местах. Пусть μ и ν — две вероятностные меры на интервале.

Теорема 14.1. *Если*

$$M \subset \left\{ \omega : \lim_{n\to\infty} \frac{\ln \nu(u_n(\omega))}{\ln \mu(u_n(\omega))} = \delta \right\}, \tag{14.8}$$

то

$$\dim_\mu M = \delta \dim_\nu M. \tag{14.9}$$

Чтобы быть уверенным в том, что отношение логарифмов в (14.8) всегда определено, условимся считать, что при $0 < \xi,\ \eta < 1$

$$\begin{aligned} \frac{\ln \xi}{\ln 0} &= \frac{\ln 1}{\ln \eta} = \frac{\ln 1}{\ln 0} = 0, \\ \frac{\ln 0}{\ln \eta} &= \frac{\ln \xi}{\ln 1} = \frac{\ln 0}{\ln 1} = \infty, \\ \frac{\ln 0}{\ln 0} &= \frac{\ln 1}{\ln 1} = 1. \end{aligned} \tag{14.10}$$

Прежде чем доказывать теорему, приведем подтверждающее ее эвристическое рассуждение и покажем, как из нее следует результат Эгглстона.

Представим себе, что для каждого ω из M отношение

$$\frac{\ln \nu(u_n(\omega))}{\ln \mu(u_n(\omega))}$$

не только сходится к δ, но что оно равно δ для всех n. Если $\{v_i\}$ — произвольное покрытие множества M цилиндрами, каждый из которых пересекается с M, то любой цилиндр имеет вид $v_i = u_n(\omega)$ при некотором n и некотором ω из M и, следовательно, $\nu(v_i) = \mu(v_i)^\delta$. Таким образом, $\sum_i \nu(v_i)^\alpha = \sum_i \mu(v_i)^{\alpha\delta}$ для любого покрытия $\{v_i\}$. Но тогда $\nu_\alpha(M) = \mu_{\alpha\delta}(M)$, откуда, очевидно, следует (14.9) Это рассуждение делает утверждение теоремы правдоподобным.

Чтобы доказать результат Эгглстона с помощью теоремы 14.1, будем считать μ мерой Лебега λ. Так как $\lambda(u_n(\omega)) = 2^{-n}$, то из теоремы следует (берем $\delta = \theta/\ln r$), что из

$$M \subset \left\{ \omega : \lim_{n\to\infty} \left[-\frac{1}{n} \ln \nu(u_n(\omega)) \right] = \theta \right\} \tag{14.11}$$

вытекает

$$\dim M = \frac{\theta}{\ln r} \dim_\nu M. \tag{14.12}$$

Если мы построим меру ν, для которой выполнено (14.11) и для которой $\nu(M) > 0$, то $\dim_\nu M = 1$ и, следовательно, в силу (14.12) $\dim M = \theta/\ln r$. Пусть ν — вероятностная мера на единичном интервале, относительно которой $\{x_n\}$ есть последовательность независимых случайных величин с $\nu\{\omega : x_n(\omega) = i\} = p_i$, $i = 0, 1, \ldots, r-1$ (см. пример 3.5). Так как

$$-\frac{1}{n} \ln \nu(u_n(\omega)) = -\sum_{i=0}^{r-1} \frac{N_i(\omega, n)}{n} \ln p_i,$$

то ясно, что (14.11) выполняется, если $M = M(p_0, \ldots, p_{r-1})$ и $\theta = -\sum_{i=0}^{r-1} p_i \ln p_i$. Так как $M(p_0, \ldots, p_{r-1})$ имеет ν-меру 1 в силу усиленного закона больших чисел, то приходим к результату Эгглстона.

В качестве следующего приложения рассмотрим число $N_{ij}(\omega, n)$ тех $k \leqslant n$, для которых $x_k(\omega) = i$ и $x_{k+1}(\omega) = j$. Если

$$\lim_{n\to\infty} \frac{1}{n} N_{ij}(\omega, n) = \pi_{ij}, \qquad i, j = 0, \ldots, r-1, \tag{14.13}$$

то (π_{ij}) есть $(r \times r)$-матрица с неотрицательными элементами, такими, что если $p_i = \sum_j \pi_{ij}$, то $\sum_j \pi_{ji} = p_i$ и $\sum_j p_i = 1$. Пусть $p_{ij} = \pi_{ij}/p_i$, тогда (p_{ij}) есть стохастическая матрица, для которой p_i — стационарные вероятности. Предположим для простоты, что p_{ij} положительны. Если ν — мера, относительно которой $\{x_1, x_2, \ldots\}$ — соответствующий марковский процесс, то

$$\lim_{n\to\infty}\left[-\frac{1}{n}\ln\nu(u_n(\omega))\right] = -\sum_{ij} p_i p_{ij}\ln p_{ij}$$

для любого ω из множества $M(\pi)$, определенного равенством (14.13). Так как $\nu(M(\pi)) = 1$, то, как и раньше, получаем

$$\dim M(\pi) = -\frac{1}{\ln r}\sum_{ij} p_i p_{ij}\ln p_{ij}.$$

Из сравнения этих результатов с теоремой Шеннона — Макмиллана — Бреймана видна их связь с энтропией.

Для третьего приложения теоремы 14.1 предположим, что $r = 3$ и $p_0 = p_2 = 1/2$, $p_1 = 0$. Пусть ν — мера, относительно которой $\{x_1, x_2, \ldots\}$ — независимые случайные величины и $\nu\{x_n = i\} = p_i$ (ν — канторова мера из § 3). Величина $-n^{-1}\ln\nu(u_n(\omega))$ равна $\ln 2$, если все $x_n(\omega)$ равны либо 0, либо 2, т. е. если ω лежит в канторовом множестве. Таким образом,

$$M \subset \left\{\omega : \lim_{n\to\infty}\left[-\frac{1}{n}\ln\nu(u_n(\omega))\right] = \ln 2\right\},$$

если M — канторово множество, так что в силу того, что из (14.11) следует (14.12),

$$\dim M = \frac{\ln 2}{\ln 3}\dim_\nu M.$$

Так как $\nu(M) = 1$, то мы получаем результат Хаусдорфа: $\dim M = \ln 2/\ln 3$.

Перейдем к доказательству теоремы 14.1. Достаточно показать, что из (14.8) вытекает неравенство

$$\dim_\mu M \geqslant \delta\dim_\nu M, \tag{14.14}$$

так как обратное неравенство получится, если поменять μ и ν местами, а δ заменить на $1/\delta$. Докажем несколько больше, а именно что неравенство (14.14) выполняется, если

$$M \subset \left\{\omega : \lim_{n\to\infty}\frac{\ln\nu(u_n(\omega))}{\ln\mu(u_n(\omega))} \geqslant \delta\right\}. \tag{14.15}$$

Сначала докажем этот результат в предположении, что $\mu(v)>0$ для любого цилиндра v, пересекающего M; позднее мы укажем на изменения, необходимые для исследования общего случая.

Достаточно показать, что если $1/\delta<\eta$ и $\dim_\mu M<\xi$, то

$$\dim_\nu M \leqslant \eta\xi.$$

Если $\omega\in M$, то в силу (14.15) $\nu(u_n(\omega))^\eta \leqslant \mu(u_n(\omega))$ для всех n, больших некоторого натурального $N(\omega)$. Пусть M_ρ — множество тех ω, которые лежат в M и при каждом n удовлетворяют одному из двух неравенств:

$$\mu(u_n(\omega)) \geqslant \rho \quad \text{или} \quad \nu(u_n(\omega))^\eta \leqslant \mu(u_n(\omega)). \tag{14.16}$$

Ясно, что M_ρ возрастает с убыванием ρ, мы покажем, что $M_\rho \uparrow M$ при $\rho \downarrow 0$. Если $\omega\in M$, возьмем $\rho=\mu(u_{N(\omega)}(\omega))>0$ (здесь нам нужно предположение, что цилиндры, пересекающие M, имеют положительную μ-меру). Если $\mu(u_n(\omega))<\rho$, то $n>N(\omega)$, так что $\nu(u_n(\omega))^\eta \leqslant \mu(u_n(\omega))$. Таким образом, ω удовлетворяет (14.16) для всех n, т. е. $\omega\in M_\rho$.

Так как $M_\rho \uparrow M$, то в силу (14.3) достаточно доказать, что

$$\dim_\nu M_\rho \leqslant \eta\xi \tag{14.17}$$

для всех положительных ρ. Предположим, что $0<\rho_1<\rho$ и $\varepsilon_1>0$. Так как $\dim_\mu M_\rho<\xi$, то

$$\sum_i \mu(v_i)^\xi < \varepsilon_1$$

для некоторого μ-ρ_1-покрытия $\{v_i\}$ множества M_ρ. Можно предположить, что все v_i пересекают M_ρ. Но тогда $v_i=u_n(\omega)$ для некоторого $\omega\in M_\rho$, так что в силу $\mu(v_i)\leqslant\rho_1<\rho$ имеем $\nu(v_i)^\eta \leqslant \mu(v_i)$. Поэтому

$$\sum_i \nu(v_i)^{\eta\xi} < \varepsilon_1.$$

Но $\{v_i\}$ есть ν-$\rho_1^{1/\eta}$-покрытие множества M_ρ, и, следовательно, так как ε_1 и ρ_1 произвольно малы, то $\nu_{\xi\eta}(M_\rho)=0$, что доказывает (14.17).

Остается снять ограничение $\mu(v)>0$ для всех цилиндров v, пересекающихся с M. Пусть E_μ — объединение всех цилиндров μ-меры 0. Так как $\dim_\mu E_\mu=0$, то $\dim_\mu A=\dim_\mu B$, если симметрическая разность множеств A и B входит в E_μ (используем (14.3)). Если ω лежит в множестве, фигурирующем в правой части соотношения (14.15) (при $\delta>0$), и если

$\mu(u_n(\omega)) = 0$ для некоторого n, то $\nu(u_m(\omega)) = 0$ для некоторого m в силу соглашения (14.10). Таким образом, $M \cap E_\mu \subset \subset M \cap E_\nu$. Мы уже знаем, что (14.14) справедливо для $M - E_\mu$, следовательно,

$$\dim_\mu M = \dim_\mu (M - E_\mu) \geqslant \delta \dim_\nu (M - E_\mu) \geqslant$$
$$\geqslant \delta \dim_\nu (M - E_\nu) = \delta \dim_\nu M.$$

Замечание. Относительно связи хаусдорфовой размерности с топологической размерностью см. Гуревич и Волмэн [1].

Гуд [1] выдвинул предположение о том, что имеет место (14.4). Доказано это соотношение было Эгглстоном [1]. Биллингслей [1, 2] получил дальнейшие результаты методами этого раздела, а Кинни и Питчер [1] применили эти методы к проблемам размерности, связанным с разложениями в непрерывные дроби. О связи с теоремой кодирования для канала без шума см. Биллингслей [3]. Реньи [2] исследовал связь между размерностью и энтропией[1]).

[1]) В работе Фюрстенберга [1] изучена связь хаусдорфовой размерности с так называемой топологической энтропией (см. Адлер и др. [1]). Соответствующие результаты близки к содержанию этого параграфа. — *Прим. ред.*

ГЛАВА 5

Кодирование

15. ТЕОРЕМА КОДИРОВАНИЯ ДЛЯ КАНАЛА БЕЗ ШУМА

Обозначения

В этой главе мы будем рассматривать одновременно несколько сдвигов и поэтому следует сразу договориться об обозначениях. Кроме конечного множества ρ (пространства состояний), понадобятся еще два множества, которые мы обозначим σ и τ. Конечное множество ρ $[\sigma, \tau]$ состоит из r $[s, t]$ элементов, причем общий элемент будем обозначать i $[j, k]$. Мы называем эти множества *алфавитами*, а их элементы — *буквами*.

Обозначим через X $[Y, Z]$ пространство бесконечных в обе стороны последовательностей $x = (\ldots, x_{-1}, x_0, x_1, \ldots)$ $[y = (\ldots, y_{-1}, y_0, y_1, \ldots), z = (\ldots, z_{-1}, z_0, z_1, \ldots)]$ элементов из ρ $[\sigma, \tau]$, а через $\mathscr{X}$ $[\mathscr{Y}, \mathscr{Z}]$ обозначим σ-поле, порожденное цилиндрами. Позволим себе некоторую вольность в обозначениях, а именно будем использовать символ x_n в двух смыслах — для обозначения n-й координаты (при фиксированном x) и для обозначения координатной переменной (при фиксированном n). Кроме того, мы будем рассматривать пространства, подобные $X \times Y$, и тогда x_n будет обозначать функцию, значение которой в точке (x, y) есть n-я координата для x. Эти соглашения относятся также к y_n и z_n. Сдвиги будем обозначать T_X, $T_{X \times Y}$ и т. д. или, если нет опасности путаницы, просто T.

Символ ρ^n $[\sigma^n, \tau^n]$ обозначает множество наборов по n элементов из ρ $[\sigma, \tau]$. Если μ — некоторая вероятностная мера на $\mathscr{X}$, которая сохраняется преобразованием T_X, обозначим всю структуру $(X, \mathscr{X}, \mu, T_X)$ просто $[X, \mu]$. Единственный меняющийся элемент здесь μ; говорят, что мера μ (или $[X, \mu]$) эргодическая [перемешивающая и т. д.], если сдвиг T_X эргодический [перемешивающий и т. д.] относительно μ. Энтропия меры $[X, \mu]$ есть энтропия сдвига T_X. Подобные замечания относятся и к пространствам Y, Z. Будем называть $[X, \mu]$ источником, или источником информации (см. § 5). Мы рассматриваем x_n как буквы, создаваемые источником, а последовательности $(x_m, \ldots, x_n)$ и $x = (\ldots, x_{-1}, x_0, x_1, \ldots)$ как *сообщения*. Мера μ описывает структуру этого источника.

Канал без шума

Теорему кодирования для канала без шума можно формулировать в терминах кодируемых сообщений как утверждение о том, как сделать эти сообщения наиболее короткими или как использовать наименьший алфавит. Мы рассмотрим теорему кодирования с этой последней точки зрения,

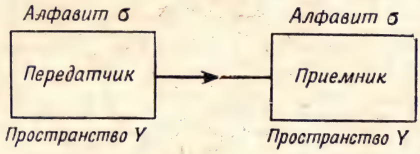


Рис. 5.

и это послужит введением к более сложным проблемам последующих параграфов.

Канал без шума состоит из передатчика и приемника, каждый из которых использует алфавит σ (рис. 5). Математически канал без шума есть просто конечный алфавит σ объема s и соответствующее пространство последовательностей $(Y, \mathscr{Y})$. (Мера здесь не участвует.) Если сообщение y

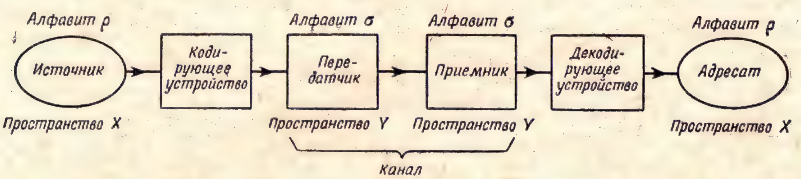


Рис. 6.

посылают по каналу, то на другом конце канала его получают без всяких искажений.

Предположим теперь, что сообщения должны быть посланы источником по каналу некоторому адресату, причем источник и адресат пользуются алфавитом ρ, а канал — алфавитом σ. Если ρ и σ различны, то, очевидно, сообщение должно быть кодировано [1]) некоторым образом до передачи и декодировано впоследствии (рис. 6).

Если источник создает, скажем, одну букву в секунду, то мы требуем, чтобы по каналу передавалась тоже одна буква в секунду — ни быстрее, ни медленнее.

[1]) Слово „код“ здесь не имеет ничего общего с секретным сообщением.

Предположим, что объем r алфавита ρ не превышает объема s алфавита σ ($r \leqslant s$). Например, ρ может состоять из $r = 10$ знаков 0, 1, ..., 9, а σ — из $s = 26$ букв A, B, ..., Z. Ясно, что сообщения, создаваемые источником, можно кодировать таким образом, чтобы они могли быть переданы по каналу (отложим на время точное математическое определение понятия кода). Действительно, возьмем некоторое взаимно однозначное отображение ψ алфавита ρ в алфавит σ; оно существует, если $r \leqslant s$. Кодируем последовательность x, заменяя каждый символ x_n символом $y_n = \psi(x_n)$, и пошлем сообщение, состоящее из этих элементов алфавита σ, по каналу. Для того чтобы восстановить посланное сообщение на другом конце канала, к полученным буквам можно применить обратное отображение ψ^{-1}.

Последовательность x может быть передана по каналу, даже если $r > s$. Это ясно, когда, например, существует такое подмножество ρ_0 множества ρ, объема не больше s, что $\mu\{x_n \in \rho_0\} = 1$. Рассмотрим менее тривиальный пример: пусть $r = 4$, $s = 2$, и пусть структура меры μ такова, что каждая буква с вероятностью 1 повторяется, т. е. $\mu\{x_{2n} = x_{2n+1}\} = 1$. (Этот источник не вполне стационарен, но пример можно изменить так, чтобы он был вполне стационарным; см. пример 3.3.) В ρ^2 имеются только четыре возможных значения (x_{2n}, x_{2n+1}). Пусть ψ — взаимно однозначное отображение этих элементов на четыре элемента из σ^2. Если сообщение кодировано с помощью ψ: $(y_{2n}, y_{2n+1}) = \psi(x_{2n}, x_{2n+1})$, то x_n можно однозначно восстановить по y_n, и, следовательно, код делает возможной передачу x по каналу.

Итак, задача состоит в том, чтобы кодировать x в y предпочтительно таким образом, чтобы затем x можно было восстановить по y. Мы должны теперь точно определить, что мы понимаем под кодом. *Код* есть измеримое отображение φ пространства X в пространство Y. Можно потребовать, чтобы код обладал некоторыми нужными свойствами.

1. *Стационарные коды*. Код φ стационарен, если равенство

$$\varphi T_X x = T_Y \varphi x \tag{15.1}$$

справедливо для всех x из X. Это означает, что структура кодирующего устройства не меняется со временем. Мы всегда будем требовать, чтобы код был стационарным; с этого момента стационарность является частью определения кода.

2. *Обратимые коды*. Код φ обратим, если он по существу взаимно однозначен, т. е. в $\mathscr{X}$ существует такое множество X_0, что $\mu(X_0) = 1$ и φ — взаимно однозначное отображение множества X_0 на $Y_0 = \varphi X_0$. Мы требуем также, чтобы $T_X X_0 = X_0$

и $\varphi A \in \mathscr{Y}$ для любого множества A, для которого $A \subset X_0$ и $A \in \mathscr{X}$ (так что, в частности, $Y_0 \in \mathscr{Y}$). Код φ называют обратимым, когда такое множество X_0 существует. Если код обратим, то посланное сообщение (по существу) однозначно восстанавливается по полученному.

В то время как понятие кода определяется безотносительно к какой бы то ни было мере μ на X, обратимость конкретного кода зависит от того, какую меру μ мы рассматриваем. Для того чтобы подчеркнуть роль меры μ, говорят, что код φ обратим *по отношению к* μ.

Предположим, что код φ обратим по отношению к μ. Покажем, что сдвиг T_X с мерой μ и сдвиг T_Y с мерой $\mu\varphi^{-1}$ изоморфны. (Значение $\mu\varphi^{-1}$ на $B \in \mathscr{Y}$ есть $\mu(\varphi^{-1}B)$; напомним, что φ всегда предполагается измеримым. Мера $\mu\varphi^{-1}$ является распределением полученных сообщений.) Очевидно, что $Y_0 \in \mathscr{Y}$, $\mu\varphi^{-1}(Y_0) = 1$ и $T_Y Y_0 = Y_0$. Пусть φ_0 — сужение φ на X_0. Предположим, что $A \subset X_0$, и положим $B = \varphi_0 A$. Тогда если $A \in \mathscr{X}$, то $B \in \mathscr{Y}$ в силу определения обратимости кода φ; если $B \in \mathscr{Y}$, то $A = X_0 \cap \varphi^{-1}B \in \mathscr{X}$ в силу измеримости φ, и если $A \in \mathscr{X}$, то, разумеется, $\mu(A) = \mu\varphi^{-1}(B)$. Так как $\varphi_0 T_X x = T_Y \varphi_0 x$ для всех x из X_0, то тройка (X_0, Y_0, φ_0) удовлетворяет условиям (I_1), (I_2) и (I_3) изоморфизма (см. § 5).

Пусть, с другой стороны, сдвиг T_X с мерой μ на X изоморфен сдвигу T_Y с мерой γ на Y. Так как T_X и T_Y обратимы, то из замечания 5, следующего за определением изоморфизма в § 5, вытекает, что они изоморфны относительно тройки (X_0, Y_0, φ_0), обладающей тем специальным свойством, что $T_X X_0 = X_0$ и $T_Y Y_0 = Y_0$. В этом случае φ_0 почти является обратимым кодом (но не вполне), так как отображает X_0 на Y_0, а не X в Y. Расширим φ_0 до отображения φ с областью определения X, полагая $\varphi x = \varphi_0 x$ для x из X_0 и взяв в качестве φx для x из $X - X_0$ некоторый элемент $(\ldots, j, j, j, \ldots)$ пространства Y, все компоненты которого одинаковы (общий образ для всех x из $X - X_0$). Тогда φ обратим по отношению к μ, его сужение на X_0 есть снова φ_0 и мера $\mu\varphi^{-1}$ идентична мере γ.

Таким образом, обратимый код по существу является изоморфизмом. Поэтому теория кодирования связана с теорией, развитой в гл. 2.

3. *Коды без предвосхищения.* Код φ называется кодом без предвосхищения, если для всех j из σ и всех целых n множество $\varphi^{-1}\{y : y_n = j\} = \{x : (\varphi x)_n = j\}$ принадлежит σ-полю, порожденному случайными величинами $\ldots, x_{n-1}, x_n$. (Обращаясь к мере μ на X, можно более общим образом потребовать только, чтобы $\varphi^{-1}\{y : y_n = j\}$ отличалось на множество

меры 0 от некоторого элемента этого σ-поля. Но мы не нуждаемся в таком расширенном определении.)

4. *Коды с обращениями без предвосхищения.* Если код φ обратим, то сужение φ_0 кода φ, которое отображает X_0 на Y_0, имеет обращение φ_0^{-1}, отображающее Y_0 на X_0. Такое обращение φ_0^{-1} можно расширить одним или несколькими способами до некоторого обратимого кода, отображающего Y в X. Можно также требовать от одного из этих расширений, отличающихся друг от друга лишь несущественно, чтобы оно являлось кодом без предвосхищения. В этом случае декодирующее устройство может не отставать от кодирующего.

Бóльшая часть этих условий накладывается на понятие кода, что затрудняет доказательство существования кодов с заданными свойствами и облегчает доказательство их несуществования.

Теоремы кодирования

В случае канала без шума проблема состоит в следующем. Для заданного источника $[X, \mu]$ с алфавитом ρ и канала без шума с алфавитом σ ищут код φ (из X в Y), обратимый по отношению к μ. Если такой код φ существует — и это является доматематической идеей, лежащей в основе проблемы, — то сообщение x может быть кодировано в виде φx и послано по каналу к приемнику, где посланное сообщение может быть восстановлено из φx. Кажется интуитивно ясным, что для каждого из двух примеров, предшествовавших определению кода, обратимый код существует. В первом из этих примеров множество ρ мало (или по крайней мере буквы, создаваемые источником, принадлежат маленькому подмножеству ρ_0 множества ρ), а во втором примере источник повторяет каждую букву дважды. Каждая из этих двух характеристик ограничивает скорость, с которой источник создает информацию. Дело в том, что обратимый код существует в том и только том случае, если энтропия источника $[X, \mu]$ не превышает $\ln s$.

Теорема 15.1. *Если энтропия h источника $[X, \mu]$ превышает $\ln s$, то не существует кода (из X в Y), обратимого по отношению к μ.*

Доказательство. Если обратимый код φ существует, то сдвиг T_X (с мерой μ) изоморфен сдвигу T_Y (с мерой $\mu\varphi^{-1}$). Но тогда T_Y также имеет энтропию h. Так как алфавит σ

содержит s букв, то (см. (7.1)) $h = h(T_Y) \leqslant \ln s$, что доказывает теорему.

Основной момент этого доказательства заключается в том, что из изоморфности T_X и T_Y мы делаем вывод, что $h(T_X) = h(T_Y)$; это указывает на важность инвариантного определения энтропии. С другой стороны, естественная мера скорости, с которой источник $[X, \mu]$ создает информацию, есть

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i_0, \ldots, i_{n-1}} \eta(\mu\{x_0 = i_0, \ldots, x_{n-1} = i_{n-1}\})$$

(т. е. $h(\mathcal{A}, T_X)$, где $\mathcal{A}$ — поле событий, наблюдаемых в момент времени 0), и если $h(T_X)$ не совпадает с этой величиной, то $h(T_X)$ не является естественной мерой скорости создания информации; это указывает на важность теоремы Колмогорова. Таким образом, доказательство фактически использует все содержание гл. 2.

По поводу обращения теоремы 15.1 мы докажем только один условный результат.

Гипотеза. *Источник* $[X, \mu]$ *имеет энтропию* h, *и сдвиг* T_X *с мерой* μ *изоморфен любому сдвигу Бернулли с энтропией* h.

Если энтропия — полный инвариант (см. § 8) для сдвигов Бернулли (соответственно перемешивающих сдвигов Маркова, сдвигов Колмогорова и т. д.), то это предположение выполнено в случае, когда источник $[X, \mu]$ является бернуллиевским (соответственно перемешивающим марковским, колмогоровским и т. д.). Возможно, что это предположение не выполняется вообще ни для какого источника.

„Теорема" 15.2. *Если* $h \leqslant \ln s$, *то при справедливости высказанного предположения существует код (из* X *в* Y*), обратимый по отношению к* μ.

Доказательство. Функция $\sum_j \eta(p_j)$ вектора вероятностей $(p_1, \ldots, p_s)$ принимает все значения между ее минимумом 0 и максимумом $\ln s$ (например, если ξ изменяется от 0 до 1, то $\sum_j \eta(\xi s^{-1} + (1-\xi)\delta_{j1})$ непрерывно меняется от 0 до $\ln s$). Поэтому если $h \leqslant \ln s$, то существуют такие вероятности $p_1, \ldots, p_s$, что $\sum_j \eta(p_j) = h$. Если γ — мера на Y, которая делает сдвиг T_Y сдвигом Бернулли $(p_1, \ldots, p_s)$, то T_Y имеет энтропию h. По нашему предположению сдвиг T_X с мерой μ изоморфен сдвигу T_Y с мерой γ; следовательно,

существует код φ, обратимый по отношению к μ (и, между прочим, обладающий свойством $\mu\varphi^{-1} = \gamma$).

Этот результат справедлив, но, быть может, бессодержателен; отсюда кавычки. Вопрос, является ли этот результат бессодержательным, эквивалентен вопросу о том, является ли энтропия полным инвариантом. Если эта „теорема" осложняется предположением, что наш код — код без предвосхищения, то это означает на самом деле, что вопросы, поставленные в § 8, осложняются требованием зависимости изоморфизма только от прошлого — сравните с результатом Синая о слабом изоморфизме.

Замечание. Начало изучению этих проблем кодирования положил Шеннон [1]. Теорема 15.1 предложена мною — по крайней мере я не встречал ее в литературе. Это, конечно, весьма простое следствие теории Колмогорова. Не встречал я и „теорему" 15.2. Теорему кодирования для каналов без шума, изложенную с других точек зрения, можно найти у Файнстейна [2], Хинчина [3] и Биллингслея [3]. В этих работах имеются варианты теоремы 15.1, по-разному неудовлетворительные, но в них содержатся конструктивные варианты „теоремы" 15.2.

16. КАНАЛ С ШУМОМ

Определения

Рассмотрим теперь канал, в котором на передаваемые сообщения действует шум, т. е. канал, в котором полученное сообщение может не быть точной копией переданного. Здесь передатчик и приемник могут использовать различные алфавиты (рис. 7).

Для математической формализации этой схемы рассмотрим пространства $(Y, \mathcal{Y})$ и $(Z, \mathfrak{Z})$. Предположим, что для каждого $y \in Y$ имеется некоторая вероятностная мера $\nu_y(\cdot)$ на $\mathfrak{Z}$. Будем считать, по техническим соображениям, что для каждого множества C из $\mathfrak{Z}$ мера $\nu_y(C)$ как функция от y измерима относительно $\mathcal{Y}$. Тройка $[Y, \nu_y, Z]$ называется *каналом*; функция $\nu_y(\cdot)$ называется *ядром* канала. Идея состоит в том, что если сообщение y передается по каналу, то полученное сообщение, которое будет, вообще говоря, искажено каким-либо случайным шумом, находится с вероятностью $\nu_y(C)$ в подмножестве C множества Z. В случае канала с шумом проблема состоит в кодировании данного

источника $[X, \mu]$ в Y так, чтобы канал использовался наиболее эффективно. Например, можно задаться вопросом, выгодно ли бороться с шумом, повторяя каждую букву дважды. (Структуру канала и источника мы предполагаем полностью известной и не рассматриваем проблему статистического вывода.)

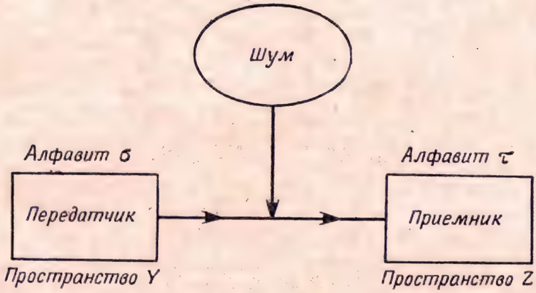


Рис. 7.

Хотя мы можем здесь без потери общности использовать один и тот же алфавит для источника, передатчика и приемника, мы все же, для упорядочения обозначений, возьмем алфавиты ρ, σ и τ соответственно.

Если $Y = Z$ и ν_y представляет собой единичную массу, сосредоточенную в y, то мы имеем канал без шума, описанный в предыдущем параграфе. Если ν_y — единичная масса для каждого y (Z может отличаться от Y), то мы имеем некоторый код из Y в Z.

Канал *стационарен*, если

$$\nu_{T_Y y}(T_Z C) = \nu_y(C).$$

Это, как обычно, означает, что свойства канала не меняются со временем. Мы всегда будем предполагать канал стационарным и включим это требование в определение.

Канал называется каналом *без предвосхищения*, если для каждого n и каждого $k \in \tau$ функция $\nu_y\{z : z_n = k\}$ измерима относительно σ-поля, порожденного элементами $\ldots, y_{n-1}, y_n$. (Код есть специальный род канала. Понятие канала без предвосхищения соответствует понятию кода без предвосхищения.)

Канал без памяти

Рассмотрим такой специальный вид канала: пусть (c_{jk}) есть $(s \times t)$-матрица неотрицательных чисел, строки и столбцы которой нумеруются элементами алфавитов σ и τ соответственно и суммы по строкам равны 1. Ядро $\nu_y(\cdot)$ задается формулой

$$\nu_y\{z_l = k_l,\ m \leqslant l \leqslant n\} = \prod_{l=m}^{n} c_{y_l k_l}.$$

Этот канал передает буквы входного сообщения независимо; если по каналу передается буква j, то буква k получается на выходе с вероятностью c_{jk}. (Разумеется, сами буквы на входе не предполагаются независимыми; в самом деле, меры на пространстве Y не участвуют в описании структуры канала.) Канал такого рода — очевидно, что он стационарный и без предвосхищения — называют *каналом без памяти* [1]). Заметим, что если $\sigma = \tau$ и $c_{jj} = 1$ для всех j, то мы имеем канал без шума, описанный в предыдущем параграфе.

Доказывая, что хорошие (в смысле выполнения некоторых свойств) коды *не существуют*, мы не будем накладывать на каналы никаких ограничений. Доказывая, что хорошие коды *существуют*, мы основное внимание уделим каналу без памяти.

Совместное распределение на входе и выходе

Предположим, что задана мера μ на Y такая, что $[Y, \mu]$ есть источник. Мы хотим определить на произведении пространств $(Y \times Z, \mathcal{Y} \times \mathcal{Z})$ некоторую вероятностную меру P, которая будет описывать совместное распределение на входе и выходе канала. Очевидно, нам нужно, чтобы для $B \in \mathcal{Y}$ и $C \in \mathcal{Z}$ выполнялось равенство

$$P(B \times C) = \int_B \nu_y(C)\, \mu(dy). \tag{16.1}$$

Этот интеграл определен в силу предположения, что $\nu_y(C)$ как функция от y измерима. Положим для множества M из $\mathcal{Y} \times \mathcal{Z}$

$$P(M) = \int_Y \nu_y\{z : (y, z) \in M\}\, \mu(dy),$$

[1]) Более общее понятие канала с конечной памятью см., например, у Хинчина [3]. (См. также Такано [1], который исправил ошибку в рассуждении Хинчина.)

при $M = B \times C$ это сводится к (16.1). Класс множеств M из $\mathcal{Y} \times \mathcal{Z}$, для которых подинтегральное выражение есть функция, измеримая относительно $\mathcal{Y}$, содержит конечные объединения непересекающихся прямоугольников $B \times C$ и в силу монотонности совпадает с $\mathcal{Y} \times \mathcal{Z}$. Итак, интеграл определен. Теперь легко проверить, что P есть вероятностная мера и что это единственная мера, удовлетворяющая (16.1).

Если источник и канал стационарны, а мы всегда предполагаем, что это так, то можно ожидать, что сдвиг $T_{Y \times Z}$ сохраняет меру P.

Теорема 16.1. *Мера P сохраняется при сдвиге $T_{Y \times Z}$*[1].

Доказательство. Достаточно показать, что $T = T_{Y \times Z}$ сохраняет меры прямоугольников $B \times C$. В силу (16.1) имеем

$$P\left(T^{-1}(B \times C)\right) = \int\limits_{T_Y^{-1}B} \nu_y\left(T_Z^{-1}C\right)\mu(dy) =$$

$$= \int\limits_{B} \nu_{T_Y^{-1}y}\left(T_Z^{-1}C\right)\mu(dy) = \int\limits_{B} \nu_y(C)\,\mu(dy) = P(B \times C).$$

Распределение, индуцированное мерой P на Y, совпадает с μ. Распределение, индуцированное на Z, есть распределение на выходе канала; как следует из теоремы 16.1, оно сохраняется при сдвиге T_Z.

Скорость передачи

Определим теперь *скорость передачи* по каналу $[Y, \nu_y, Z]$ для источника $[Y, \mu]$. Каждому из трех сдвигов $T_{Y \times Z}$ (с мерой P), T_Y [с мерой $\mu(B) = P(B \times Z)$] и T_Z [с мерой $P(Y \times C)$] приписывается некоторая энтропия. Скорость определяется формулой

$$R = h(T_Y) + h(T_Z) - h(T_{Y \times Z}).$$

Интуитивный смысл этой формулы мы обсудим несколько позже.

Более детально скорость можно изучить с помощью теоремы Колмогорова. Пусть $\mathcal{B}$ (соответственно $\mathcal{C}$) обозначает конечное поле в $\mathcal{Y} \times \mathcal{Z}$ с атомами $\{(y, z) : y_0 = j\}$, $j \in \sigma$ (соответственно $\{(y, z) : z_0 = k\}$, $k \in \tau$). Обозначим через $\mathcal{B} \vee \mathcal{C}$ поле событий из $Y \times Z$, наблюдаемых в момент

[1]) $T_{Y \times Z}$ действует по формуле $T_{Y \times Z}(y, z) = (T_Y y, T_Z z)$. — *Прим. ред.*

времени 0. Нетрудно показать, что если $\mathscr{B}_0$ — конечное поле в $\mathscr{Y}$ с атомами $\{y : y_0 = j\}$, $j \in \sigma$, то $H\left(\bigvee_{l=0}^{n-1} T_Y^{-l}\mathscr{B}_0\right) = H\left(\bigvee_{l=0}^{n-1} T^{-l}\mathscr{B}\right)$, где $T = T_{Y\times Z}$. Поэтому в силу теоремы Колмогорова $h(T_Y) = h(T_Y, \mathscr{B}_0) = h(T, \mathscr{B})$. Применяя аналогичное рассуждение к сдвигу T_Z, видим, что если ввести обозначения

$$H\left(\bigvee_{k=0}^{n-1} T^{-k}\mathscr{B}\right) = H(y_0, \ldots, y_{n-1}),$$

$$H\left(\bigvee_{k=0}^{n-1} T^{-k}\mathscr{C}\right) = H(z_0, \ldots, z_{n-1}),$$

$$H\left(\bigvee_{k=0}^{n-1} T^{-k}(\mathscr{B} \vee \mathscr{C})\right) = H(y_0, \ldots, y_{n-1}, z_0, \ldots, z_{n-1}),$$

то

$$R = \lim_{n\to\infty} \frac{1}{n}\{H(y_0, \ldots, y_{n-1}) + H(z_0, \ldots, z_{n-1}) - H(y_0, \ldots, y_{n-1}, z_0, \ldots, z_{n-1})\}. \quad (16.2)$$

Используя свойство (A_1') из § 6, можно переписать выражение, стоящее в фигурных скобках в (16.2), двумя способами. Имеем:

$$\begin{aligned} H(y_0, \ldots, y_{n-1}) + H(z_0, \ldots, z_{n-1}) - H(y_0, \ldots, y_{n-1}, z_0, \ldots, z_{n-1}) &= \\ = H(y_0, \ldots, y_{n-1}) - H(y_0, \ldots, y_{n-1} \mid z_0, \ldots, z_{n-1}) &= \\ = H(z_0, \ldots, z_{n-1}) - H(z_0, \ldots, z_{n-1} \mid y_0, \ldots, y_{n-1}). & \quad (16.3) \end{aligned}$$

Средний член формулы (16.3) имеет следующую интерпретацию. Если бы мы получили сообщение $y_0, \ldots, y_{n-1}$, то (среднее) количество полученной информации равнялось бы $H(y_0, \ldots, y_{n-1})$ — прирост информации равен устраненной неопределенности. Однако если сообщение $y_0, \ldots, y_{n-1}$ было передано по каналу таким образом, что мы получили вместо него искаженное сообщение $z_0, \ldots, z_{n-1}$, то к той неопределенности, которой обладало посланное сообщение $y_0, \ldots, y_{n-1}$, прибавляется еще некоторая неопределенность — (среднее) количество этой неопределенности есть $H(y_0, \ldots, y_{n-1} \mid z_0, \ldots, z_{n-1})$. Поэтому разность между этими двумя величинами, т. е. средний член формулы (16.3), является разумной мерой (среднего) количества получаемой информации. Итак,

полученная информация =
= информация в посланном сообщении
минус неопределенность от искажения при передаче.

Таким образом,

$$R = \lim_{n\to\infty} \frac{1}{n}\{H(y_0, \ldots, y_{n-1}) - H(y_0, \ldots, y_{n-1} \mid z_0, \ldots, z_{n-1})\}$$

представляет собой количество полученной информации, приходящейся *на букву*, и, следовательно, задает *скорость*. Величина

$$\lim_{n\to\infty} \frac{1}{n} H(y_0, \ldots, y_{n-1} \mid z_0, \ldots, z_{n-1}) = h(T_Y) - R = h(T_{Y\times Z}) - h(T_Z)$$

называется *ненадежностью* (на букву). Если энтропия источника фиксирована, то увеличение скорости передачи равносильно уменьшению ненадежности.

Если мы принимаем ту интерпретацию энтропии и условной энтропии, которая была предложена в гл. 2, то эти определения естественны: высокая скорость (малая ненадежность) хороша, а низкая скорость (большая ненадежность) плоха. Другой подход к измерению эффективности пары источник — канал состоит в рассмотрении проблемы реального построения эффективного алгоритма для восстановления посланного сообщения по полученному. Хотя этот второй подход косвенным образом опирается на материал § 18 и 19, мы не будем изучать его систематически [1]).

Явно вычислить скорость мы можем только в специальных случаях. Пусть (c_{jk}) есть $(s \times t)$-матрица, описывающая канал без памяти. Так как

$$H(z_0, \ldots, z_{n-1} \mid y_0, \ldots, y_{n-1}) =$$
$$= -\int_Y \left\{ \sum_{k_0, \ldots, k_{n-1}} (c_{y_0k_0} \ldots c_{y_{n-1}k_{n-1}}) \ln (c_{y_0k_0} \ldots c_{y_{n-1}k_{n-1}}) \right\} \mu(dy),$$

то

$$H(z_0, \ldots, z_{n-1} \mid y_0, \ldots, y_{n-1}) = nH(z_0 \mid y_0) \tag{16.4}$$

независимо от распределения величины y_n. Для любого множества вероятностей $p_1, \ldots, p_s$ на алфавите σ определим другое множество вероятностей $q_1, \ldots, q_t$ на алфавите τ по формуле

$$q_k = \sum_j p_j c_{jk}$$

---

[1]) Сравнение этих двух подходов обсуждается у Файнстейна [2, гл. 3].

и положим

$$r(p_1, \ldots, p_s) = \sum_j \eta(p_j) - \sum_k q_k \sum_j \eta\left(\frac{p_j c_{jk}}{q_k}\right) =$$

$$= \sum_k \eta(q_k) - \sum_j p_j \sum_k \eta(c_{jk}) =$$

$$= \sum_j \eta(p_j) + \sum_k \eta(q_k) - \sum_{jk} \eta(p_j c_{jk}). \quad (16.5)$$

Если

$$\mu\{y : y_0 = j\} = p_j, \qquad j \in \sigma, \quad (16.6)$$

то, даже если y_n не независимы относительно μ,

$$r(p_1, \ldots, p_s) = H(y_0) - H(y_0 \mid z_0) =$$

$$= H(z_0) - H(z_0 \mid y_0) =$$

$$= H(y_0) + H(z_0) - H(y_0, z_0). \quad (16.7)$$

Если выполняется (16.6) и *если y_n независимы*, то

$$H(y_0, \ldots, y_{n-1}) + H(z_0, \ldots, z_{n-1}) - H(y_0, \ldots, y_{n-1}, z_0, \ldots, z_{n-1}) =$$

$$= nH(y_0) + nH(z_0) - nH(y_0, z_0) = nr(p_1, \ldots, p_s),$$

так что скорость задается формулой

$$R = r(p_1, \ldots, p_s). \quad (16.8)$$

Это соотношение может не выполняться, если y_n не независимы.

Пример 16.1. Пусть имеется канал без памяти, где $s = 2$, $t = 3$, и матрица (c_{jk}) есть

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

Для любых вероятностей на входе (p_1, p_2) матрица $(p_j c_{jk}/q_k)$ есть

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix},$$

так что $H(y_0 \mid z_0) = 0$. Даже если y_n не независимы, то в силу свойств (A_4) и (A_3) из § 6

$$H(y_0, \ldots, y_{n-1} \mid z_0, \ldots, z_{n-1}) \leqslant \sum_{i=0}^{n-1} H(y_i \mid z_0, \ldots, z_{n-1}) \leqslant$$
$$\leqslant \sum_{i=0}^{n-1} H(y_i \mid z_i) = 0,$$

так что ненадежность исчезает. Для этого канала скорость всегда совпадает с энтропией на входе[1]).

Пример 16.2. Пусть имеется канал без памяти, где все строки матрицы (c_{jk}) одинаковы: c_{jk} не зависят от j. Тогда $H(z_0 \mid y_0) = H(z_0)$ и, следовательно, в силу (16.4)

$$H(z_0, \ldots, z_{n-1}) - H(z_0, \ldots, z_{n-1} \mid y_0, \ldots, y_{n-1}) =$$
$$= H(z_0, \ldots, z_{n-1}) - nH(z_0).$$

В силу свойства (A_4) из § 6 эта величина равна нулю. Скорость всегда нуль, независимо от входа; по этому каналу не может быть передано никакой информации[2]).

Пример, описывающий промежуточную ситуацию между двумя крайними, представленными примерами 16.1 и 16.2, показывает, что скорость обычно зависит не только от источника и не только от канала, а от их взаимодействия.

Пример 16.3. Пусть имеется канал без памяти, где $s = t = 4$, и матрица (c_{jk}) есть

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

Из (16.5) следует

$$r(p_1, \ldots, p_4) = \eta(p_1) + \eta(p_2) + \eta(p_3 + p_4). \qquad (16.9)$$

Предположим теперь, что y_n независимы относительно μ, так что (16.9) задает скорость передачи, а

$$\eta(p_1) + \eta(p_2) + \eta(p_3) + \eta(p_4) \qquad (16.10)$$

[1]) Канал, подобный этому, называется каналом *без потерь*. См. Файнстейн [2].

[2]) Характеристика каналов без памяти, обладающих этим свойством, дана у Файнстейна [2].

— энтропия источника $[Y, \mu]$. Если $p_1 = p_2 = 0$, то скорость (16.9) равна нулю, в то время как энтропия (16.10) может достигать $\ln 2$; такой источник плохо согласуется с каналом. С другой стороны, если $p_4 = 0$, то скорость (16.9) и энтропия (16.10) совпадают, хотя их общее значение может равняться нулю; такой источник хорошо согласуется с каналом.

Пропускная способность канала

Скорость передачи зависит и от источника $[Y, \mu]$, и от канала $[Y, \nu_y, Z]$; для фиксированного канала $[Y, \nu_y, Z]$ скорость есть функция R_μ меры μ, описывающей вероятностную структуру источника. Верхнюю грань этих скоростей

$$C = \sup R_\mu \tag{16.11}$$

называют *пропускной способностью.*

Пропускная способность — это скорость, с которой информация может быть передана по каналу, если он используется наиболее эффективным образом — путем выбора источника, „подходящего“ этому каналу. (В действительности мы обычно не знаем, достигается ли эта верхняя грань.) По техническим соображениям мы должны различать две пропускные способности. Если верхняя грань в (16.11) берется по всем мерам μ, которые сохраняются сдвигом T_Y, то мы имеем *стационарную пропускную способность* C_s. Если верхняя грань берется только по тем μ, относительно которых сдвиг T_Y эргодичен, то мы имеем *эргодическую пропускную способность* C_e. Очевидно, что $C_e \leqslant C_s$.

Заметим, что для канала без шума, описанного в предыдущем параграфе, $C_e = C_s = \ln s$. Вообще вычислять пропускную способность довольно трудно, но для канала без памяти это сделать можно.

Теорема 16.2. *Для канала без памяти*

$$C_s = C_e = \max r(p_1, \ldots, p_s),$$

где $(p_1, \ldots, p_s)$ *пробегают все возможные распределения вероятностей на алфавите* σ.

Доказательство. Выбирая $p_1, \ldots, p_s$ так, чтобы максимизировать $r(p_1, \ldots, p_s)$ (используя компактность), и вспоминая, что сдвиг Бернулли эргодичен, в силу (16.8) получаем $C_e \geqslant \max r(p_1, \ldots, p_s)$.

Для доказательства неравенства $C_s \leqslant \max r(p_1, \ldots, p_s)$ достаточно показать, что для любой меры μ на входе и для любого n справедливо неравенство

$$H(z_0, \ldots, z_{n-1}) - H(z_0, \ldots, z_{n-1} \mid y_0, \ldots, y_{n-1}) \leqslant \\ \leqslant nH(z_0) - nH(z_0 \mid y_0).$$

Очевидно, что

$$H(z_0, \ldots, z_{n-1}) \leqslant H(z_0) + \ldots + H(z_{n-1}) = nH(z_0),$$

и наше утверждение следует из (16.4).

Это рассуждение показывает, что для канала без памяти верхняя грань в определениях C_s и C_e достигается; для более сложных каналов это доказать трудно.

В примере 16.1 пропускная способность канала равна $\ln s = \ln 2$; в примере 16.2 она равна 0, а в примере 16.3 она равна $\ln 3$.

Эргодичность процесса передачи

Можно задаться вопросом, какие условия на канал $[Y, \nu_y, Z]$ обеспечивают эргодичность входной — выходной меры P [определенной соотношением (16.1)], какова бы ни была мера μ. Необходимые и достаточные условия неизвестны. Сформулируем и докажем следующий результат, который в дальнейшем использоваться не будет.

Теорема 16.3. *Канал без памяти обладает тем свойством, что мера P эргодична, какова бы ни была мера μ*

Доказательство. Достаточно показать (теорема 1.4), что

$$\lim_{n\to\infty} \frac{1}{n} \sum_{l=0}^{n-1} P\left(E \cap T^{-l}E'\right) = P(E)\, P\left(E'\right) \tag{16.12}$$

для цилиндров E и E'. Будем доказывать (16.12) в предположении, что $E = \{(y, z)\colon\ y_0 = j,\ z_0 = k\}$ и $E' = \{(y, z)\colon\ y_0 = j',\ z_0 = k'\}$; более общий случай доказывается аналогично. Левая часть равенства (16.12) есть

$$\lim_{n\to\infty} \frac{1}{n} \sum_{l=0}^{n-1} P\{y_0 = j,\ z_0 = k,\ y_l = j',\ z_l = k'\} =$$

$$= \lim_{n\to\infty} \frac{1}{n} \sum_{l=0}^{n-1} \mu\left\{y_0 = j,\ y_l = j'\right\} c_{jk} c_{j'k'} =$$

$$= \mu\left\{y_0 = j\right\} \mu\left\{y_0 = j'\right\} c_{jk} c_{j'k'} = P(E)\, P(E'),$$

где предпоследнее равенство справедливо в силу эргодичности μ.

Замечание. Изложенные здесь идеи принадлежат Шеннону [1]. Макмиллан [1], Файнстейн [1] и Хинчин [3] дали им точную математическую формулировку.

Брейман [2] обобщил теорему 16.2, доказав, что $C_e = C_s$ и верхняя грань достигается для некоторого класса каналов с конечной памятью. По поводу обобщения теоремы 16.3 см. Адлер [1].

17. ТЕОРЕМА КОДИРОВАНИЯ ДЛЯ КАНАЛА С ШУМОМ

Проблема

В общих словах теорему кодирования для канала с шумом можно описать следующим образом. Сообщения посылаются от источника $[X, \mu]$ по данному каналу $[Y, \nu_y, Z]$ к адресату (рис. 8). Мы должны вставить в схему кодирующее

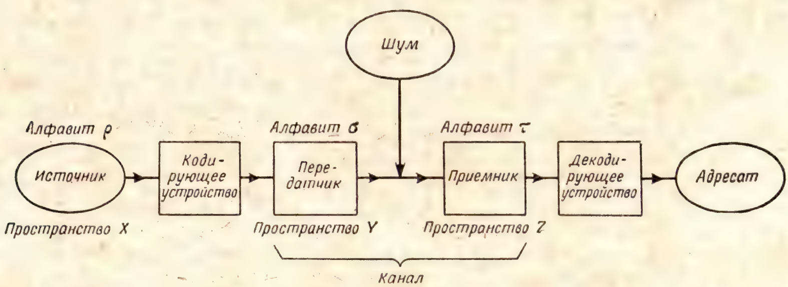


Рис. 8.

устройство, которое будет работать эффективно. Эффективность измеряют скоростью, с которой информация передается от источника к приемнику, игнорируя при этом вопрос о том, как использовать эту информацию в приемнике для восстановления посланного сообщения.

Код φ из пространства X в пространство Y приводит к сложному каналу из X в Z, т. е. сообщение x посылается в кодирующее устройство, после чего закодированное сообщение y посылается по каналу и на выходе получается сообщение z. Пусть энтропия источника равна h и пропускная способность — скажем, стационарная пропускная способность — канала равна h. Прямая теорема кодирования утверждает, что при $h \leqslant C$ существует такой код φ, что ско-

рость передачи по сложному каналу равна h. Обратная теорема кодирования утверждает, что при $h > C$ такого кода φ не существует. Есть различные варианты этих теорем.

Пусть нам дан канал $[Y, \nu_y, Z]$ и код φ из X в Y. Очевидно, что $\nu_{\varphi(x)}(\cdot)$ — ядро и, следовательно, $[X, \nu_{\varphi(x)}(\cdot), Z]$ — сложный канал. Рассмотрим для заданной меры μ на X меру P на $X \times Y \times Z$, определенную формулой

$$P(A \times B \times C) = \int\limits_{A \cap \varphi^{-1}B} \nu_{\varphi(x)}(C)\,\mu(dx). \qquad (17.1)$$

Мера P индуцирует частные распределения P_X, $P_{X \times Y}$ и т. д. Заметим, что $P_X = \mu$, $P_Y = P_X\varphi^{-1} = \mu\varphi^{-1}$ [так как $P(A \times B \times C) = P(A \cap \varphi^{-1}B \times Y \times C)$] и $P_{X \times Z}$ задает совместное распределение на входе и выходе для источника $[X, \mu]$ и сложного канала $[X, \nu_{\varphi(x)}, Z]$.

Пусть h — энтропия источника, C_s — стационарная пропускная способность канала $[Y, \nu_y, Z]$. Скорость передачи по сложному каналу $[X, \nu_{\varphi(x)}, Z]$ (с мерами P_X, P_Z, $P_{X \times Z}$) для источника $[X, \mu]$ есть

$$R_\mu = h(T_X) + h(T_Z) - h(T_{X \times Z}).$$

Ненадежность есть величина

$$h(T_X) - R_\mu = h(T_{X \times Z}) - h(T_Z),$$

на которую эта скорость меньше энтропии источника $[X, \mu]$. Мы ставим перед собой задачу разыскания кода, при котором скорость высока (ненадежность мала), игнорируя при этом, как уже отмечалось, проблему, заключающуюся в восстановлении x по z.

Простое обращение

Следующий результат является частным случаем теоремы 17.3.

Теорема 17.1. *Если код φ обратим по отношению к μ, то $R_\mu \leqslant C_s$. Следовательно, при $h > C_s$ не существует кода, обратимого по отношению к μ, для которого $R_\mu = h$.*

Доказательство. По определению стационарной пропускной способности скорость передачи по каналу $[Y, \nu_y, Z]$ для источника $[Y, \mu\varphi^{-1}]$ есть величина

$$h(T_Y) + h(T_Z) - h(T_{Y \times Z}),$$

не превышающая C_s.

Следовательно, достаточно доказать равенство

$$h(T_X) + h(T_Z) - h(T_{X \times Z}) = h(T_Y) + h(T_Z) - h(T_{Y \times Z}).$$

(Здесь сдвигу T_Z соответствует мера P_Z.) Так как по предположению код φ обратим, то сдвиги T_X и T_Y изоморфны и, следовательно, $h(T_X) = h(T_Y)$. Таким образом, нужно только доказать, что

$$h(T_{X \times Z}) = h(T_{Y \times Z}).$$

Но если T_X изоморфен T_Y относительно тройки (X_0, Y_0, φ_0), то, как легко видеть, $T_{X \times Z}$ изоморфен $T_{Y \times Z}$ относительно тройки $(X_0 \times Z, Y_0 \times Z, \psi_0)$, где $\psi_0(x, z) = (\varphi_0 x, z)$, и теорема доказана.

В случае когда $[Y, \nu_y, Z]$ — канал без шума, теорема 17.1 сводится к теореме 15.1. И снова для доказательства необходимо инвариантное определение энтропии.

Если мы предполагаем, что μ эргодична, то и $\mu\varphi^{-1}$ эргодична, и в этой теореме можно заменить C_s на C_e. Заметим, что мы не предполагали, что φ — код без предвосхищения.

На практике проблема декодирования заключается в восстановлении x по z, а не по y, ибо предположение обратимости в теореме 17.1 неестественно. В теореме 17.3 это предположение снимается.

Комментарии к прямой теореме для канала без памяти

Нам представляется соблазнительной попытка доказать для канала без памяти условный результат, подобный „теореме“ 15.2. Наше предположение таково:

Гипотеза. *Источник* $[X, \mu]$ *имеет энтропию* h, *и сдвиг* T_X *с мерой* μ *изоморфен любому сдвигу Бернулли с энтропией* h.

Ясно сознавая, что нам не удастся дать для нижеследующей „теоремы“ ни условного, ни безусловного доказательства, мы все же сформулируем утверждение и исследуем возникающие весьма поучительные трудности.

Напомним, что структура канала без памяти задается $(s \times t)$-матрицей (c_{jk}), описанной в § 16, и пропускная способность $C = C_e = C_s$ дается максимумом величины (16.5).

„Теорема“ 17.2. *Пусть имеется канал без памяти, источник* $[X, \mu]$ *удовлетворяет высказанной гипотезе*

и $h \leqslant C$. *Тогда существует такой обратимый по отношению к* μ *код* φ, *что скорость передачи* R_μ *по сложному каналу равна* h.

Рассмотрим сначала специальный канал без памяти $[Y, \nu_y, Z]$, для которого можно провести доказательство, — канал из примера 16.3. Пропускная способность этого канала равна $\ln 3$. Если $h \leqslant \ln 3$, то для $p_4 = 0$ и подходящим образом выбранных p_1, p_2 и p_3 величины (16.9) и (16.10) принимают одинаковое значение h. Пусть γ — мера на пространстве Y, которая делает сдвиг T_Y сдвигом Бернулли $(p_1, \ldots, p_4)$. Тогда $h(T_Y) = h$, скорость передачи по каналу $[Y, \nu_y, Z]$ для источника $[Y, \gamma]$, равна h и ненадежность исчезает. По предположению существует код φ (из X в Y), обратимый по отношению к μ, такой, что $\gamma = \mu\varphi^{-1}$. Из обратимости кода φ следует, как и в доказательстве теоремы 17.1, что скорость передачи по сложному каналу для источника $[X, \mu]$ имеет то же значение h, что скорость передачи по каналу $[Y, \nu_y, Z]$ для источника $[Y, \mu\varphi^{-1}]$.

Это рассуждение работает, если только канал обладает тем специальным свойством, что при $h \leqslant C$ существует на Y мера γ, удовлетворяющая условиям:

1) T_Y — сдвиг Бернулли относительно меры γ;

2) $[Y, \gamma]$ имеет энтропию h;

3) скорость передачи по каналу для источника $[Y, \gamma]$ равна h (или, что эквивалентно в силу условия 2, ненадежность равна 0).

Этим свойством обладает канал из примера 16.3 и любой канал без шума („теорема“ 15.2), но им не обладает общий канал без памяти — например, в случае, когда все c_{jk} строго положительны.

Для общего канала без памяти можно доказать, что при $h \leqslant C$ существует мера γ на пространстве Y, удовлетворяющая условиям 2 и 3. Однако для того, чтобы провести наше доказательство полностью, необходимо при невыполнении условия 1 усилить основную гипотезу таким образом, чтобы обеспечить изоморфизм сдвигов T_X (с мерой μ) и T_Y (с мерой γ). Для того чтобы превратить „теорему“ 15.2 в настоящую теорему, скажем, для перемешивающего марковского источника $[X, \mu]$, нужно потребовать полноты энтропии для перемешивающих сдвигов Маркова. Для того чтобы превратить „теорему“ 17.2 в настоящую теорему для того же источника, нужно потребовать полноты энтропии для некоторого более широкого класса сдвигов — достаточно широкого, чтобы включить сдвиг T_Y с мерой γ.

Эти замечания предлагают некоторую возможную программу доказательства прямой теоремы кодирования для канала без памяти. Параграфы 18 и 19 можно рассматривать как конечное конструктивное приближение к этой программе: мы построим меру γ, приближенно удовлетворяющую условиям 2 и 3, и затем построим почти обратимый код φ, переводящий μ в γ.

Усиление обращения *

Теорема 17.1 показывает, что если между источником и передатчиком действует обратимый код, то скорость передачи от источника к приемнику не может превышать пропускной способности канала. Интуитивно ясно, что это утверждение должно быть справедливо и в случае необратимого кода, и в случае *случайного кода*. Случайный код — это просто канал $[X, \zeta_x, Y]$; если его ядро $\zeta_x(\cdot)$ есть единичная масса в точке φx, то мы имеем снова обычный код φ. Сложный канал, который получается в результате введения случайного кода, — это канал $[X, \beta_x, Z]$, где ядро $\beta_x(\cdot)$ определяется формулой

$$\beta_x(C) = \int_Y \zeta_x(dy)\, \nu_y(C).$$

(Заметим, что $\beta_x(C)$ — измеримая функция от x.) Мы покажем, что скорость передачи по каналу $[X, \beta_x, Z]$ для источника $[X, \mu]$ не может превышать скорости передачи по каналу $[Y, \nu_y, Z]$ для источника $[Y, \lambda]$, где $\lambda(B) = \int_X \mu(dx)\, \zeta_x(B)$; разумеется, это вторая скорость не может превышать стационарной пропускной способности C_s канала $[Y, \nu_y, Z]$.

Определим меру P на $X \times Y \times Z$ формулой

$$P(A \times B \times C) = \int_A \int_B \int_C \mu(dx)\, \zeta_x(dy)\, \nu_y(dz) = \int_A \int_B \mu(dx)\, \zeta_x(dy)\, \nu_y(C). \tag{17.2}$$

[Если код $[X, \zeta_x, Y]$ не случайный, то (17.2) сводится к (17.1).] Пусть $\mathcal{A}$ — поле событий, наблюдавшихся в момент времени 0, для пространства X, вложенного в произведение $X \times Y \times Z$, т. е. $\mathcal{A}$ — конечное подполе поля $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ с атомами $\{(x, y, z) : x_0 = i\}$, $i \in \rho$. Аналогично пусть $\mathcal{B}$ и $\mathcal{C}$ — поля событий, наблюдавшихся в момент времени 0, соответственно для

X и Z, вложенных в $X \times Y \times Z$. Если T — сдвиг на $X \times Y \times Z$ (с мерой P, определенной формулой (17.2)), то две упоминавшиеся выше скорости есть

$$R(x \to z) = h(\mathcal{A}, T) + h(\mathcal{C}, T) - h(\mathcal{A} \vee \mathcal{C}, T)$$

и

$$R(y \to z) = h(\mathcal{B}, T) + h(\mathcal{C}, T) - h(\mathcal{B} \vee \mathcal{C}, T).$$

Мы хотим доказать интуитивно ясное соотношение $R(x \to z) \leqslant \leqslant R(y \to z)$, которое сводится к

$$h(\mathcal{A} \vee \mathcal{C}, T) - h(\mathcal{A}, T) \geqslant h(\mathcal{B} \vee \mathcal{C}, T) - h(\mathcal{B}, T). \quad (17.3)$$

Пусть $\mathcal{X}_0 = \bigvee_{n=-\infty}^{\infty} T^n \mathcal{A}, \mathcal{Y}_0 = \bigvee_{n=-\infty}^{\infty} T^n \mathcal{B}$ и $\mathcal{Z}_0 = \bigvee_{n=-\infty}^{\infty} T^n \mathcal{C}$; заметим, что, например, $\mathcal{X}_0$ состоит из множеств вида $A \times Y \times Z$, где A принадлежит $\mathcal{X}$. Основной факт, который мы используем, заключается в том, что если $M \in \mathcal{Z}_0$, то

$$P\{M \,\|\, \mathcal{X}_0 \vee \mathcal{Y}_0\} = P\{M \,\|\, \mathcal{Y}_0\} \text{ п. в.} \quad (17.4)$$

Это соотношение, аналогичное свойству марковости, вытекает из (17.2): если $M = X \times Y \times C$, то общее значение правой и левой частей равенства (17.4) в точке (x, y, z) есть $\nu_y(C)$.

В силу (12.5) имеем $h(\mathcal{A} \vee \mathcal{C}, T) = H(\mathcal{A} \vee \mathcal{C} \mid \mathcal{A}^- \vee \mathcal{C}^-)$, где $\mathcal{A}^- = \bigvee_{n<0} T^n \mathcal{A}$ и $\mathcal{C}^- = \bigvee_{n<0} T^n \mathcal{C}$. Тогда по теореме 12.3

$$h(\mathcal{A} \vee \mathcal{C}, T) = H(\mathcal{A} \vee \mathcal{C} \mid \mathcal{A}^- \vee \mathcal{C}^-) = H(\mathcal{C} \mid \mathcal{C}^- \vee \mathcal{X}_0) + H(\mathcal{A} \mid \mathcal{A}^-).$$

Аналогично

$$(\mathcal{B} \vee \mathcal{C}, T) = H(\mathcal{B} \vee \mathcal{C} \mid \mathcal{B}^- \vee \mathcal{C}^-) = H(\mathcal{C} \mid \mathcal{C}^- \vee \mathcal{Y}_0) + H(\mathcal{B} \mid B^-).$$

Наконец, $h(\mathcal{A}, T) = H(\mathcal{A} \mid \mathcal{A}^-)$ и $h(\mathcal{B}, T) = H(\mathcal{B} \mid \mathcal{B}^-)$. Поэтому (17.3) эквивалентно неравенству

$$H(\mathcal{C} \mid \mathcal{C}^- \vee \mathcal{X}_0) \geqslant H(\mathcal{C} \mid \mathcal{C}^- \vee \mathcal{Y}_0).$$

Так как в силу свойства (C_3) из § 12

$$H(\mathcal{C} \mid \mathcal{C}^- \vee \mathcal{X}_0) \geqslant H(\mathcal{C} \mid \mathcal{C}^- \vee \mathcal{X}_0 \vee \mathcal{Y}_0),$$

то вся задача сводится к доказательству равенства

$$H(\mathcal{C} \mid \mathcal{C}^- \vee \mathcal{X}_0 \vee \mathcal{Y}_0) = H(\mathcal{C} \mid \mathcal{C}^- \vee \mathcal{Y}_0). \quad (17.5)$$

Если мы покажем, что

$$H(\mathcal{C} \mid \mathcal{C}^{-n} \vee \mathcal{X}_0 \vee \mathcal{Y}_0) = H(\mathcal{C} \mid \mathcal{C}^{-n} \vee \mathcal{Y}_0), \quad (17.6)$$

где $\mathscr{C}^{-n} = \bigvee\limits_{-n \leqslant k < 0} T^k\mathscr{C}$, то справедливость соотношения (17.5) будет следовать из теоремы 12.1. Но в силу свойства (C_1) из § 12 соотношение (17.6) представляется в виде

$$H(\mathscr{C} \vee \mathscr{C}^{-n} \mid \mathscr{X}_0 \vee \mathscr{Y}_0) - H(\mathscr{C}^{-n} \mid \mathscr{X}_0 \vee \mathscr{Y}_0) = \\ = H(\mathscr{C} \vee \mathscr{C}^{-n} \mid \mathscr{Y}_0) - H(\mathscr{C}^{-n} \mid \mathscr{Y}_0).$$

И, наконец, в силу (17.4) имеем

$$H(\mathscr{C}^{-n} \mid \mathscr{X}_0 \vee \mathscr{Y}_0) = H(\mathscr{C}^{-n} \mid \mathscr{Y}_0)$$

и

$$H(\mathscr{C} \vee \mathscr{C}^{-n} \mid \mathscr{X}_0 \vee \mathscr{Y}_0) = H(\mathscr{C} \vee \mathscr{C}^{-n} \mid \mathscr{Y}_0).$$

Мы доказали следующий результат.

Теорема 17.3. *Для любого случайного кода $R(x \to z) \leqslant R(y \to z) \leqslant C_s$. Если h—энтропия источника $[X, \mu]$—превышает C_s, то не существует случайного кода, для которого $R(x \to z) = h$.*

Замечание. Теоремы 17.1 и 17.3 по существу содержатся в результатах Пинскера [1].

§ 18. ТЕОРЕМА ФАЙНСТЕЙНА

Вернемся теперь к анализу прямой теоремы кодирования для канала с шумом. Ограничимся рассмотрением канала без памяти. Сама теорема приводится в § 19. Здесь мы докажем два предварительных результата, принадлежащих Файнстейну.

Решающая схема

Канал без памяти задается $(s \times t)$-матрицей (c_{jk}) из § 16. Его пропускная способность $C = C_s = C_e$ есть максимум выражения (16.5).

Теорема 18.1. *Рассмотрим канал без памяти с пропускной способностью C. Если $0 < \varepsilon < C$, то существует такое положительное целое число $b_1(\varepsilon)$, что σ^b при $b \geqslant b_1(\varepsilon)$ состоит из N различных точек $u_1, \ldots, u_N$, а τ^b распадается на N непересекающихся множеств $V_1, \ldots, V_N$, где u_l и V_l таковы, что*

$$P\{(z_1, \ldots, z_b) \in V_l \mid (y_1, \ldots, y_b) = u_l\} > 1 - \varepsilon, \quad l = 1, \ldots, N, \tag{18.1}$$

и

$$N > e^{b(C-\varepsilon)}. \tag{18.2}$$

Если $u_l = (j_1, \ldots, j_b) \in \sigma^b$, то левая часть неравенства (18.1) есть просто

$$\sum c_{j_1 k_1} \ldots c_{j_b k_b},$$

где сумма берется по наборам $(k_1, \ldots, k_b)$ из подмножества V_l множества τ^b. Следовательно, условные вероятности определяются только каналом, так что теорема применима к любой мере γ на входе Y.

Сущность этой теоремы состоит в следующем. Допустим, посланное сообщение $(y_1, \ldots, y_b)$ получено в виде сообщения $(z_1, \ldots, z_b)$. Если известно, что $(y_1, \ldots, y_b)$ — один из наборов $u_1, \ldots, u_N$, а $(z_1, \ldots, z_b)$ принадлежит V_l, то в силу (18.1) представляется разумным предположить, что $(y_1, \ldots, y_b)$ есть действительно набор u_l. И если источник может кодироваться обратимым или почти обратимым образом, причем так, что переданное сообщение с большой вероятностью принадлежит $\{u_1, \ldots, u_N\}$, то посланное источником сообщение должно с хорошей точностью восстанавливаться по полученному, и, следовательно, ненадежность должна быть малой. Наборы u_l и V_l дают схему для решения вопроса о том, каково было переданное сообщение. (См. обсуждение скорости и ненадежности в § 16.) Эта идея лежит в основе теоремы кодирования следующего параграфа.

Идея доказательства теоремы 18.1 такова. Предположим, что алфавит σ содержит ровно две буквы, скажем, 0 и 1. Пусть b — четное число, $b = 2a$, и пусть последовательность u_1 состоит из a нулей, за которыми следует a единиц, а последовательность u_2 состоит из a единиц, за которыми следует a нулей. Если b велико и $(y_1, \ldots, y_b) = u_1$, то с высокой условной вероятностью частоты в $(z_1, \ldots, z_a)$ почти совпадают с первой строкой матрицы (c_{jk}), в то время как частоты в $(z_{a+1}, \ldots, z_b)$ почти совпадают со второй строкой матрицы. Если $(y_1, \ldots, y_b) = u_2$, то верно то же самое, но две первые строки матрицы (c_{jk}) меняются местами. Так как эти две строки различны (если только пропускная способность не равна 0), то распределение $(z_1, \ldots, z_b)$ при условии $(y_1, \ldots, y_b) = u_1$ сильно отличается (при большом b) от распределения при условии $(y_1, \ldots, y_b) = u_2$ в силу того, что нули и единицы расположены в u_1 и u_2 не одинаковым образом; и это позволяет нам найти непересекающиеся множества V_1 и V_2, для которых имеет место (18.1). Доказательство опирается на тщательный анализ условных распределений на выходе для *большого* класса различных входных последовательностей.

Так как заключение теоремы 18.1 тем сильнее, чем меньше ε, то, доказывая ее, можно предполагать, что

$$0<\varepsilon<\frac{1}{2}.$$

Поскольку результат не зависит от того, какова входная мера γ, можно выбрать ее так, чтобы упростить доказательство. Выберем меру γ так, чтобы y_n были независимы относительно γ, причем $\gamma\{y_0=j\}=p_j$, где $p_1, \ldots, p_s$ выбраны так, чтобы функция $r(p_1, \ldots, p_s)$, определенная формулой (16.5), была максимизирована. Тогда в силу теоремы 16.2

$$C=H(O)-H(O\,|\,I),$$

где $q_k=\sum_j p_j c_{jk}$, величина

$$H(O)=\sum_k \eta(q_k)$$

— энтропия на одну букву на выходе канала, а

$$H(O\,|\,I)=\sum_{jk} p_j \eta(c_{jk})$$

— условная энтропия на выходе при фиксированном входе. (Везде в доказательстве индекс j пробегает алфавит σ, а индекс k пробегает алфавит τ.)

Выберем λ так, что

$$\lambda>1, \quad \frac{s}{\lambda^2}<\frac{\varepsilon}{2}, \tag{18.3}$$

где s — объем алфавита σ. Далее, выберем K так, что

$$-3s\lambda^{3/2}\sum_k \ln q_k<K \tag{18.4}$$

и

$$-3\lambda^{3/2}\sum_{jk} \ln c_{jk}<K, \tag{18.5}$$

где суммы берутся только по тем k, для которых $q_k>0$, и по тем (j, k), для которых $c_{jk}>0$. Наконец, выберем положительное целое число $b_1(\varepsilon)$ так, что при $b \geqslant b_1(\varepsilon)$ (что мы в дальнейшем будем предполагать)

$$\frac{\ln(4/\varepsilon)+2K\sqrt{b}}{b}<\varepsilon. \tag{18.6}$$

Обозначим символом Π меру на $\sigma^b\times\tau^b$, индуцированную набором $(y_1, \ldots, y_b, z_1, \ldots, z_b)$:

$$\Pi(E)=P\{(y_1, \ldots, y_b, z_1, \ldots, z_b)\in E\}, \quad E\subset\sigma^b\times\tau^b. \tag{18.7}$$

Мера Π приписывает точке $(j_1, \ldots, j_b, k_1, \ldots, k_b)$ массу $p_{j_1} \ldots p_{j_b} c_{j_1 k_1} \ldots c_{j_b k_b}$. Если $B \subset \sigma^b\ [c \subset \tau^b]$, будем писать $\Pi(B)\,[\Pi(C)]$ вместо $\Pi(B \times \tau^b)\,[\Pi(\sigma^b \times C)]$; это не должно привести к путанице. Наконец, для $C \subset \tau^b$ и $u \in \sigma^b$ мы пишем

$$\Pi_u(C) = P\{(z_1, \ldots, z_b) \in C \mid (y_1, \ldots, y_b) = u\}. \qquad (18.8)$$

Если $u = (j_1, \ldots, j_b)$, то очевидно, что $\Pi_u(C)$ есть $\sum c_{j_1 k_1} \ldots c_{j_b k_b}$, где суммирование ведется по всем $(k_1, \ldots, k_b)$ из C.

Для $u \in \sigma^b$ и $j \in \sigma$ пусть $N(j \mid u)$ обозначает число компонент набора u, равных j; аналогично для $v \in \tau^b$ и $k \in \tau$ пусть $N(k \mid v)$ обозначает число компонент набора v, равных k; наконец, для $u \in \sigma^b$, $v \in \tau^b$, $j \in \sigma$ и $k \in \tau$ пусть $N(jk \mid uv)$ есть число тех l, $1 \leqslant l \leqslant b$, для которых l-е компоненты наборов u и v равны соответственно j и k. Тогда

$$\Pi(v) = \exp \sum_k N(k \mid v) \ln q_k \qquad (18.9)$$

и

$$\Pi_u(v) = \exp \sum_{jk} N(jk \mid uv) \ln c_{jk}. \qquad (18.10)$$

Говорят, что u есть p-последовательность, если

$$|N(j \mid u) - b p_j| < \lambda \sqrt{b}$$

для всех $j \in \sigma$; говорят, что последовательность v порождена последовательностью u, если

$$|N(jk \mid uv) - N(j \mid u) c_{jk}| < \lambda \sqrt{N(j \mid u)}$$

для всех $j \in \sigma$ и $k \in \tau$.

Нам потребуются следующие четыре леммы, из которых первые две непосредственно вытекают из неравенства Чебышева и условий (18.3).

Лемма 1. *Если S — множество p-последовательностей, то*

$$\Pi(S) > 1 - \frac{\varepsilon}{2} \geqslant \frac{1}{2}.$$

Лемма 2. *Если G_u — множество последовательностей из τ^b, порожденных последовательностью u, то*

$$\Pi_u(G_u) > 1 - \frac{\varepsilon}{2}.$$

Лемма 3. *Если последовательность v порождена некоторой p-последовательностью, то*

$$\Pi(v) \leqslant e^{-bH(O)+K\sqrt{b}}.$$

Лемма 4. *Если B_u — число элементов множества G_u, то для каждой p-последовательности u*

$$B_u \leqslant e^{bH(O\mid I)+K\sqrt{b}}.$$

Для доказательства леммы 3 заметим, что если u порождает v, то

$$N(k|v) = \sum_j N(jk|uv) \geqslant \sum_j \left[N(j|u)\,c_{jk} - \lambda\sqrt{N(j|u)}\right],$$

в то время как если u есть p-последовательность, то

$$bp_j - \lambda\sqrt{b} \leqslant N(j|u) \leqslant bp_j + \lambda\sqrt{b} \leqslant 2\lambda b,$$

так что

$$N(k|v) \geqslant b\sum_j p_j c_{jk} - s\lambda\sqrt{b} - 2s\lambda^{3/2}\sqrt{b} \geqslant bq_k - 3s\lambda^{3/2}\sqrt{b}.$$

Отсюда и из (18.9) и (18.4) следует лемма 3.

Для доказательства леммы 4 заметим, что если u есть p-последовательность и она порождает v, то

$$\begin{aligned} N(jk|uv) &\leqslant N(j|u)\,c_{jk} + \lambda\sqrt{N(j|u)} \leqslant \\ &\leqslant \left(bp_j + \lambda\sqrt{b}\right)c_{jk} + \lambda\sqrt{bp_j + \lambda\sqrt{b}} \leqslant \\ &\leqslant bp_j c_{jk} + 3\lambda^{3/2}\sqrt{b}, \end{aligned}$$

так что в силу (18.10) и (18.5)

$$\Pi_u(v) \geqslant e^{-bH(O\mid I)-K\sqrt{b}}.$$

Так как $\Pi_u(G_u) \leqslant 1$, то лемма доказана.

Для доказательства самой теоремы 18.1 рассмотрим элементы $u_1, \ldots, u_N$ из σ^b и подмножества $V_1, \ldots, V_N$ из τ^b, удовлетворяющие следующим четырем свойствам:

1) каждая u_l есть p-последовательность;

2) $\Pi_{u_l}(V_l) > 1 - \varepsilon$ для каждого l;

3) каждое подмножество V_l состоит из последовательностей, порожденных последовательностью u_l и не содержащихся в $V_1 \cup \ldots \cup V_{l-1}$, т. е. $V_l = G_{u_l} - (V_1 \cup \ldots \cup V_{l-1})$;

4) не существует u_{N+1} и V_{N+1} таких, что $u_1, \ldots, u_{N+1}$ и $V_1, \ldots, V_{N+1}$ удовлетворяют условиям 1, 2 и 3.

В силу лемм 1 и 2 такие u_l и V_l существуют (быть может, при $N=1$). Затем если u есть p-последовательность, не принадлежащая множеству $u_1, \ldots, u_N$, то

$$\Pi_u(G_u \cap (V_1 \cup \ldots \cup V_N)) \geqslant \frac{\varepsilon}{2}. \qquad (18.11)$$

В противном случае мы могли бы взять $u_{N+1} = u$ и

$$V_{N+1} = G_u - (V_1 \cup \ldots \cup V_N),$$

противоречащие условию 4, так как по лемме 2 $\Pi_{u_{N+1}}(V_{N+1}) = \Pi_u(G_u) - \Pi_u(G_u \cap (V_1 \cup \ldots \cup V_N)) > 1 - \varepsilon/2 - \varepsilon/2 = 1 - \varepsilon$. С другой стороны, если u принадлежит множеству $u_1, \ldots, u_N$, скажем, $u = u_l$, то $\Pi_u(G_u \cap (V_1 \cup \ldots \cup V_N)) \geqslant \Pi_u(V_l) = \Pi u_l(V_l) > 1 - \varepsilon > \varepsilon/2$. Таким образом, неравенство (18.11) выполняется для любой p-последовательности u.

Поэтому

$$\Pi(V_1 \cup \ldots \cup V_N) \geqslant \sum_{u \in S} \Pi(u) \Pi_u(V_1 \cup \ldots V_N) \geqslant \frac{\varepsilon}{2} \Pi(S) \geqslant \frac{\varepsilon}{4}$$

в силу леммы 1. Из леммы 3 следует, что число элементов в объединении $V_1 \cup \ldots \cup V_N$ не меньше

$$\frac{\varepsilon}{4} e^{bH(O) - K\sqrt{b}}.$$

Но в силу леммы 4 число элементов $V_1 \cup \ldots \cup V_N$ не больше

$$Ne^{bH(O \mid I) + K\sqrt{b}}.$$

Комбинируя эти оценки и используя (18.6), получаем

$$N \geqslant \frac{\varepsilon}{4} e^{bC - 2K\sqrt{b}} > e^{b(C-\varepsilon)},$$

чем доказательство завершается. (Если множества V_l не исчерпывают всего τ^b, дополним одно из них.)

Применения

Вспоминая еще раз, что теорема 18.1 справедлива при любой входной мере γ на пространстве Y, предположим, что для входной меры

$$P\{(y_1, \ldots, y_b) \in \{u_1, \ldots, u_N\}\} = \\ = \gamma\{(y_1, \ldots, y_b) \in \{u_1, \ldots, u_N\}\} = 1. \qquad (18.12)$$

Наш ближайший результат состоит в том, что в такой обстановке переданному сообщению $(y_1, \ldots, y_b)$ отвечает лишь малая неопределенность, если полученное сообщение $(z_1, \ldots, z_b)$ известно. Остается проблема конструирования почти обратимого кода, для которого свойство (18.12) выполняется.

Теорема 18.2. *Для* u_l *и* V_l *из теоремы* 18.1 *справедливо неравенство*

$$H(y_1, \ldots, y_b | z_1, \ldots, z_b) < \eta(\varepsilon) + \eta(1-\varepsilon) + \varepsilon \ln t^b \qquad (18.13)$$

при условии, что выполняется свойство (18.12) *и* $\varepsilon < 1/e$.

Доказательство. Вспомним, что объем алфавита τ равен t. Мы должны ограничить сверху величину

$$H(y_1, \ldots, y_b | z_1, \ldots, z_b) = H(\mathscr{B}' | \mathscr{C}'),$$

где $\mathscr{B}'$ — конечное подполе поля $\mathscr{Y} \times \mathscr{Z}$ с атомами $\{(y, z) : (y_1, \ldots, y_b) = u\}$, $u \in \sigma^b$, а $\mathscr{C}'$ — конечное подполе с атомами $\{(y, z) : (z_1, \ldots, z_b) = v\}$, $v \in \tau^b$. Пусть $\mathscr{B}$ имеет атомы $B_l = \{(y, z) : (y_1, \ldots, y_b) = u_l\}$, $l = 1, \ldots, N$. На самом деле эти множества исчерпывают $Y \times Z$ с точностью до некоторого множества меры 0; объединим это множество меры 0 с одним из множеств B_l (это не внесет изменений в вычисления). Наконец, пусть $\mathscr{C}$ имеет атомы $C_l = \{(y, z) : (z_1, \ldots, z_b) \in V_l\}$, $l = 1, \ldots, N$.

В силу (18.12) каждый нетривиальный атом поля $\mathscr{B}'$ отличается на множество меры 0 от некоторого нетривиального атома поля $\mathscr{B}$, и обратно. Отсюда и из того факта, что $\mathscr{C} \subset \mathscr{C}'$, получаем

$$H(y_1, \ldots, y_b | z_1, \ldots, z_b) = H(\mathscr{B}' | \mathscr{C}') = H(\mathscr{B} | \mathscr{C}') \leqslant H(\mathscr{B} | \mathscr{C}). \qquad (18.14)$$

Для любого l в силу (18.1) имеем $P(C_l | B_l) > 1 - \varepsilon$. Из этого следует, что $\sum_l P(B_l) P(C_l^c \mid B_l) < \varepsilon$, и по теореме 6.2

$$H(\mathscr{B} | \mathscr{C}) \leqslant \eta(\varepsilon) + \eta(1-\varepsilon) + \varepsilon \ln (N-1), \qquad (18.15)$$

при условии, что ε так мало, что функция $\eta(t)$ не убывает на $[0, \varepsilon]$ и не возрастает на $[1-\varepsilon, 1]$, что справедливо при $\varepsilon < 1/e$. Для того чтобы получить (18.13), воспользуемся теперь неравенствами (18.14) и (18.15).

Замечание. Файнстейн [1, 2] первый доказал результаты этого параграфа (для каналов с конечной памятью). Приведенное здесь доказательство теоремы 18.1 содержится у Вольфовица [1]. Помимо обобщений теоремы 18.1, детальное исследование Вольфовица содержит результаты, обратные этой теореме, гласящие, что если (18.1) справедливо, то N не может превышать некоторой верхней грани. С помощью этих результатов Вольфовиц получил обратную теорему кодирования.

19. БЛОКОВЫЕ КОДЫ

Определение

Рассмотрим канал из предыдущего параграфа — канал без памяти — вместе с *эргодическим* источником $[X, \mu]$ с энтропией h. Цель настоящего параграфа — показать, что если $h < C$ и $\delta > 0$, то существует такой код φ без предвосхищения (из X в Y), что если сообщение x посылается по сложному каналу, то скорость превышает $h - \delta$.

Мы докажем теорему в терминах *b-блоковых кодов*. Пусть для любого n

$$\bar{x}_n = (x_{nb+1}, \ldots, (x_{nb+b});$$

тогда $\bar{x} = (\ldots, \bar{x}_{-1}, \bar{x}_0, \bar{x}_1, \ldots)$ есть элемент пространства $\overline{X}$ последовательностей элементов из ρ^b. Аналогично определим $\bar{y}$, $\overline{Y}$, $\bar{z}$, $\overline{Z}$. Под b-блоковым кодом φ мы понимаем отображение φ: $X \to Y$, определяемое (измеримым и стационарным) кодом $\bar{\varphi}$: $\overline{X} \to \overline{Y}$ с помощью формулы

$$((\varphi x)_{nb+1}, \ldots, \varphi x)_{nb+b}) = (\bar{\varphi}\bar{x})_n.$$

Заметим, что b-блоковый код измерим, но не вполне стационарен, будучи стационарным только для блоков длины b, т. е. $\varphi T_X^b x = T_Y^b \varphi x$. Если $\bar{\varphi}$ — код без предвосхищения, то b-блоковый код φ также называют кодом без предвосхищения. (В этом случае передатчик должен запаздывать по времени на b единиц по сравнению с источником. Если даны $\ldots, x_{nb-1}, x_{nb}$, т. е. известны $\ldots, \bar{x}_{n-2}, \bar{x}_{n-1}$, то кодирующее устройство может выдать блок $\bar{y}_{n-1}$ кодированного сообщения и буквы $y_{(n-1)b+1}, \ldots, y_{nb}$ этого сообщения могут быть переданы в течение следующих b единиц времени.) Код, который мы получим, зависит только от настоящего в том смысле, что $(\bar{\varphi}\bar{x})_n$ зависит только от $\bar{x}_n$.

Так как код φ нестационарен, предыдущие определения скорости, ненадежности и т. д. непосредственно неприменимы. Мы определим эти величины для x, y, z, разделив на b соответствующие величины для $\bar{x}$, $\bar{y}$, $\bar{z}$, хотя нетрудно видеть, что такие пределы, как $\lim\limits_n n^{-1} H(z_1, \ldots, z_n)$, существуют в любом случае.

Прямая теорема в терминах блоковых кодов

Теорема 19.1. *Рассмотрим канал без памяти и некоторый эргодический источник $[X, \mu]$. Пусть C — пропускная способность канала и h — энтропия источника. Если $h < C$*

и $\delta > 0$, *то для некоторого* b *существует такой* b-*блоковый код* φ, *что скорость передачи сообщения* x *по сложному каналу превосходит* $h - \delta$.

Доказательство. Выберем ε так, что

$$h + \varepsilon < C - \varepsilon,$$

$$\eta(\varepsilon) + \eta(1 - \varepsilon) + \varepsilon \ln t < \frac{\delta}{2},$$

$$\varepsilon \ln r < \frac{\delta}{2},$$

$$\varepsilon < \frac{1}{e}.$$

(Как всегда, r и t — объемы алфавитов ρ и τ.) Все вычисления основываются на вероятностях

$$P(A \times B \times C) = \int\limits_{A \cap \varphi^{-1} B} \nu_{\varphi(x)}(C)\, \mu(dx),$$

где ν_y — ядро рассматриваемого канала и φ — тот блоковый код, который нужно построить.

Если b превосходит $b_0(\varepsilon)$ из теоремы 13.2, то независимо от того, каков код φ, алфавит ρ^b распадается на два множества H и L так, что

$$P\{(x_1, \ldots, x_b) \in L\} < \varepsilon$$

и

$$P\{(x_1, \ldots, x_b) = w\} > e^{-b(h+\varepsilon)} \tag{19.1}$$

для любого $w \in H$. Если b превосходит $b_1(\varepsilon)$ из теорем 18.1 и 18.2, то

$$H(y_1, \ldots, y_b | z_1, \ldots, z_b) < \eta(\varepsilon) + \eta(1 - \varepsilon) + \varepsilon b \ln t, \tag{19.2}$$

если только код φ таков, что $(y_1, \ldots, y_b)$ с вероятностью 1 принадлежит множеству элементов $u_1, \ldots, u_N$ из σ^b. Напомним (см. (18.2)), что

$$N \geqslant e^{b(C-\varepsilon)}.$$

Выберем и зафиксируем некоторое число b, превышающее и $b_0(\varepsilon)$, и $b_1(\varepsilon)$. Далее, из (19.1) следует, что число элементов множества H меньше, чем $e^{b(h+\varepsilon)} < e^{b(C-\varepsilon)} \leqslant N$. Поэтому существует отображение ψ, переводящее ρ^b в σ^b так, что множество H отображается взаимно однозначно на собственное подмножество $\psi(H)$ множества $\{u_1, \ldots, u_N\}$ и все элементы множества L отображаются на некоторый эле-

мент u_l, не принадлежащий $\psi(H)$. Определим $\bar{\varphi}$ (а следовательно, и φ) формулой

$$(\bar{\varphi}\bar{x})_n = \psi(\bar{x}_n).$$

По построению $\bar{y}_1 = \psi(\bar{x}_1)$ принадлежит множеству $\{u_1, \ldots, u_N\}$, так что в силу (19.2)

$$H(\bar{y}_1 | \bar{z}_1) < \eta(\varepsilon) + \eta(1-\varepsilon) + \varepsilon b \ln t < \frac{b\delta}{2}. \tag{19.3}$$

Если $\bar{y}_1$ принадлежит множеству $\psi(H)$, то тем самым $\bar{x}_1$ полностью определен; следовательно, если $u \in \psi(H)$, то это означает, что величина

$$\sum_{w \in \rho^b} \eta(P\{\bar{x}_1 = w | \bar{y}_1 = u\}) \tag{19.4}$$

обращается в нуль. Так как ρ^b содержит r^b элементов, то (19.4) во всяком случае не больше $\ln r^b$. Из соотношения $P\{\bar{y}_1 \in \psi(H)\} = P\{\bar{x}_1 \in H\}$ следует

$$H(\bar{x}_1 | \bar{y}_1) \leqslant \varepsilon \ln r^b \leqslant \frac{b\delta}{2}. \tag{19.5}$$

Далее,

$$H(\bar{x}_1 | \bar{z}_1) \leqslant H(\bar{x}_1, \bar{y}_1 | \bar{z}_1) = H(\bar{y}_1 | \bar{z}_1) + H(\bar{x}_1 | \bar{y}_1, \bar{z}_1) \leqslant$$
$$\leqslant H(\bar{y}_1 | \bar{z}_1) + H(\bar{x}_1 | \bar{y}_1).$$

Применяя теперь (19.3) и (19.5), получаем

$$H(\bar{x}_1 | \bar{z}_1) < b\delta.$$

Но

$$H(\bar{x}_1, \ldots, \bar{x}_n | \bar{z}_1, \ldots, \bar{z}_n) \leqslant \sum_{i=1}^{n} H(\bar{x}_i | \bar{z}_1, \ldots, \bar{z}_n) \leqslant$$
$$\leqslant \sum_{i=1}^{n} H(\bar{x}_i | \bar{z}_i) = nH(\bar{x}_1 | \bar{z}_1),$$

так что

$$\frac{1}{n} H(\bar{x}_1, \ldots, \bar{x}_n | \bar{z}_1, \ldots, \bar{z}_n) < b\delta.$$

Переходя к пределу, мы видим, что ненадежность для b-блокового процесса ($\bar{x}$ переходит в $\bar{z}$ по сложному каналу) меньше чем $b\delta$. Разделив на b, получаем, что ненадежность при передаче самого x меньше чем δ. Поэтому скорость превосходит $h - \delta$.

Заметим, что построенный код φ, вообще говоря, необратим. Так или иначе это несущественно, ибо проблема декодирования состоит в возвращении от z к x, а не от y

к x. Мы достигли нашей цели, состоявшей в оценке ненадежности сложного канала величиной δ.

На протяжении этой главы мы измеряли эффективность кода, связывающего источник с каналом, посредством скорости передачи информации от источника к приемнику. В § 17 мы игнорировали проблему использования информации в приемнике для восстановления посланного сообщения. С другой стороны, в теореме 19.1 мы теоретически строим эффективный код, явным образом сталкиваясь с указанной проблемой и используя при этом теорему Файнстейна. Однако реальное построение кодов в практической ситуации производится совершенно иначе.

Очевидно, что теоремы 17.1 и 17.3 о несуществовании некоторых кодов могут быть также сформулированы в терминах b-блоковых кодов. С другой стороны, не ясно, как переформулировать теорему 19.1 таким образом, чтобы обойтись без блоковых кодов.

Замечание. Эти теоремы принадлежат Шеннону [1], хотя его доказательства были неполными. Файнстейн [1] дал первое полное доказательство. См. также Файнстейн [1], Хинчин [3], Такано [1], Добрушин [1] и Пинскер [1]. Более широкие исследования в теории связи см. у Реза [1] и Мейер-Эпплера [1].

Литература[1])

Абрамов Л. М.

1. Энтропия производного автоморфизма, *ДАН СССР*, **128**, № 4 (1959), 647—650.
2. Об энтропии потока, *ДАН СССР*, 128, № 5 (1959), 873—875.
3. Энтропия автоморфизма соленоидальной группы, *Теория вероятн. и ее примен.*, **4**, вып. 3 (1959), 249—254.
4. Некоторые вопросы метрической теории динамических систем, Диссертация, МГУ, 1959.

Абрамов Л. М., Рохлин В. А.

1. Энтропия косого произведения преобразований с инвариантной мерой, *Вестник ЛГУ, сер. мат. и мех.*, **7**, вып. 2 (1962), 5—13.

Адлер (Adler R. L.)

1. Ergodic and mixing properties of infinite memory channels, *Proc. Amer. Math. Soc.*, **12** (1961), 924—930.
2. On a conjecture of Fomin, *Proc. Amer. Math. Soc.*, **13** (1962), 433—436.
3. A note on the entropy of skew product transformations. *Proc. Amer. Math. Soc.*, **14** (1963), 665—669.

Адлер, Конхейм, Мак-Эндрью (Adler R. L., Konheim A. G., McAndrew M. H.)

1. Topological entropy, *Trans. Amer. Math. Soc.*, **114** (1965), 309—319.

Биллингслей (Billingsley P.)

1. Hausdorff dimension in probability theory, *Ill. J. Math.*, **4** (1960), 187—209.
2. Hausdorff dimension in probability theory. II, *Ill. J. Math.*, **5** (1961), 291—298.
3. On the coding theorem for the noiseless channel, *Ann. Math. Statist.*, **32** (1961), 594—601.

Биркгоф (Birkhoff G. D.)

1. Proof of the ergodic theorem., *Proc. Nat. Acad. USA*, **17** (1931), 656—660.

Блюм, Хансон (Blum J. R., Hanson D. L.)

1. On the isomorphism problem for Bernoulli schemes, *Bull. Amer. Math. Soc.*, **69** (1963), 221—223.

---

[1]) Литература, отмеченная звездочкой, добавлена при переводе. — *Прим. ред.*

Браун (Brown T. A.)

1. Entropy and conjugacy, *Ann. Math. Statist.*, **34** (1963), 226—232.

Брейман (Breiman L.)

1. The individual ergodic theorem of information theory, *Ann. Math. Statist.*, 28 (1957), 809—811. Поправка к статье в *Ann. Math. Statist.*, **31** (1960), 809—810.
2. On achieving channel capacity in finite-memory channels, *Ill. J. Math.*, 4 (1960), 246—252.

Вольфовиц Дж.

1. Теоремы кодирования теории информации, «Мир», М., 1967.

Генис А. Л.

1. Метрические свойства эндоморфизмов n-мерного тора, *ДАН СССР*, **138**, № 5 (1961), 991—993.

Грейвс (Graves R.)

1. The Greek myths, v. I, II, Penguin Books, 1955.

Гуд (Good I. J.)

1. The fractional dimensional theory of continued fractions, *Proc. Camb. Phil. Soc.*, **37** (1941), 199—228.

Гуревич Б. М.

1. Энтропия потока орициклов, *ДАН СССР*, **136**, № 4 (1960), 768—770.

Гуревич В., Волмэн Г.

1. Теория размерности, ИЛ, М., 1948.

Данфорд Н., Шварц Дж. Т.

1. Линейные операторы. Общая теория, ИЛ, М., 1962.

Дёблин (Doeblin W.)

1. Remarques sur la théorie métrique des fractions continues, *Compositio Math.*, **7** (1940), 353—371.

Добрушин Р. Л.

1. Общая формулировка основной теоремы Шеннона в теории информации, *УМН*, **14**, вып. 6 (1959), 3—104.

Дуб Дж. Л.

1. Вероятностные процессы, ИЛ, М., 1956.

Зигмунд А.

1. Тригонометрические ряды, т. II, «Мир», М., 1965.

Кац М.

1. Статистическая независимость в теории вероятностей, анализе и теории чисел, ИЛ, М., 1963.

Кинни, Питчер (Kinney J. R., Pitcher T. S.)

1. The dimension of some sets defined in terms of f-expansions, *Z. Wahr.*, **4**, № 4 (1966), 293—315.

Колмогоров А. Н.

1. Основные понятия теории вероятностей, М., 1936. (Впервые на немецком языке, 1933).
2. Новый метрический инвариант транзитивных динамических систем и автоморфизмов пространств Лебега, *ДАН СССР*, **119**, № 5 (1958), 861—864.
3. Об энтропии на единицу времени как метрическом инварианте автоморфизмов, *ДАН СССР*, **124**, № 4 (1959), 754—755.

Кузьмин Р. О.

1. Об одной задаче Гаусса, *ДАН СССР*, *сер. А* (1928), 375—380.

Леви (Lévy P.)

1. Sur les lois de probabilité dont dépendent les quotients complets et incomplets d'une fraction continue, *Bull. Soc. Math.*, **57** (1929), 178—194.
2. Théorie de l'addition des variables aléatoires, Paris, 1937.

Макмиллан (McMillan B.)
1. The basic theorems of information theory, *Ann. Math. Stat.*, **24**, (1953), 196—219.

Мейер-Эпплер (Meyer-Eppler W.)
1. Grundlagen und Anwendungen der Informationstheorie, Kommunikation und Kybernetik in Einzeldarstellungen, 1, Springer, 1959.

Мешалкин Л. Д.
1. Один случай изоморфизма схем Бернулли, *ДАН СССР*, **128**, № 1 (1959), 41—44.

Нейман (von Neumann J.)
1. Proof of the quasiergodic hypothesis, *Proc. Nat. Acad. USA*, **18** (1932), 70—82.
2. Einige Sätze über messbare Abbildungen, *Ann. Math.*, **33** (1932), 574—586.

Пинскер М. С.
1. Информация и информационная устойчивость случайных величин и процессов, Изд-во АН СССР, М., 1960.
2. Динамические системы с вполне положительной и нулевой энтропией, *ДАН СССР*, **133**, № 5 (1960), 1025—1026.

Райт (редактор) (Wright F. B.)
1. Ergodic theory, New York, 1963.

Реза (Reza T. M.)
1. An introduction to information theory, New York, 1961.

Реньи (Rényi A.)
1. Representations of real numbers and their ergodic properties, *Acta Math. Acad. Sci. Hungar.*, 8 (1957), 477—493.
2. Dimension, entropy and information, Transactions of the Second Prague Conference on Information Theory, Statistical Decision Functions, Random Processes, 1959, стр. 545—556.

Рисс (Riesz F.)
1. Sur la théorie ergodique, *Comm. Math. Helvetici*, **17** (1945), 221—239.

Розенблат (Rosenblatt M.)
1. Random processes, New York, 1962.

Рохлин В. А.
1. Об энтропии метрического автоморфизма, *ДАН СССР*, **124**, № 5 (1959), 980—983.
2. Новый прогресс в теории преобразований с инвариантной мерой, *УМН*, **15**, вып. 4 (1960), 3—26.
3. Точные эндоморфизмы пространства Лебега, *Изв. АН СССР, сер. мат.*, **25**, № 4 (1961), 499—530.
4. Об энтропии автоморфизма компактной коммутативной группы, *Теория вероятн. и ее примен.*, **6**, № 3 (1961), 351—352.
5. Аксиоматическое определение энтропии преобразования с инвариантной мерой, *ДАН СССР*, **148**, № 4 (1963), 779—781.
*6. Избранные вопросы метрической теории динамических систем, *УМН*, **4**, вып. 2 (1949), 57—125.
*7. Об основных понятиях теории меры, *Матем. сб.*, **25(67)**, № 1 (1949), 107—150.

Рохлин В. А., Синай Я. Г.
1. Построение и свойства инвариантных измеримых разбиений, *ДАН СССР*, **141**, № 5 (1961), 1038—1041.

Рыль-Нарджевский (Ryll-Nardzewski C.)
1. On the ergodic theorems. II, *Studia Math.*, **12** (1951), 74—79.

Синай Я. Г.
1. О понятии энтропии динамической системы, *ДАН СССР*, **124**, № 4 (1959), 768—771.

2. О потоках с конечной энтропией, *ДАН СССР*, **125**, № 6 (1959), 1200—1202.
3. Динамические системы и стационарные марковские процессы, *Теория вероятн. и ее примен.*, **5**, № 3 (1960), 335—338.
4. Геодезические потоки на многообразиях отрицательной постоянной кривизны, *ДАН СССР*, **131**, № 4 (1960), 752—755.
5. Геодезические потоки на компактных поверхностях отрицательной кривизны, *ДАН СССР*, **136**, № 3 (1961), 549—552.
6. Динамические системы со счетнократным лебеговским спектром, *Изв. АН СССР, сер. мат.*, **25**, № 6 (1961), 899—924.
7. Вероятностные идеи в эргодической теории, Int. Congress of Math. (1962), 540—559.
8. Слабый изоморфизм преобразований с инвариантной мерой, *ДАН СССР*, **147**, № 4 (1962), 797—800.

*9. О слабом изоморфизме преобразований с инвариантной мерой, *Матем. сб.*, **63(105)**, № 1 (1963), 23—42.

Такано (Takano K.)
1. On the basic theorems of information theory, *Ann. Inst. Stat. Math.*, **9** (1958), 53—77.

Томасян (Thomasian A. J.)
1. An elementary proof of the AEP of information theory, *Ann. Math. Statist.*, **31** (1960), 452—456.

Успенский (Uspenski J. V.)
1. Introduction to mathematical probability, New York, 1937.

Файнстейн А. (Feinstein A.)
1. A new basic theorem of information theory, *IRE, Trans. PGIT*, Sept. 1954, 2—22.
2. Основы теории информации, ИЛ, М., 1960.

Феллер В.
1. Введение в теорию вероятностей и ее приложения, т. I, II, «Мир», М., 1967.

Фюрстенберг (Furstenberg H.)

*1. Disjointness in ergodic theory, minimal sets and a problem in Diophantine approximation, *Math. Syst. Theory*, **1**, № 1 (1967), 1—50.

Халмош П. (Halmos P. R.)
1. Measurable transformations, *Bull. Amer. Math. Soc.*, **55** (1949), 1015—1034.
2. Теория меры, ИЛ, М., 1953.
3. Лекции по эргодической теории, ИЛ, М., 1959.
4. Entropy in ergodic theory, мимеографические записи, The University of Chicago, 1959.
5. Новый прогресс в эргодической теории, *сб. Математика*, **6 : 3** (1962), 17—27.

Харди, Райт (Hardy G. H., Wright E. M.)
1. An introduction to the theory of numbers, Oxford, 1959.

Харрис (Harris T. E.)
1. On chains of infinite order, *Pacific J. Math.*, **5** (1955), 707—724.

Хинчин А. Я.
1. Metrische Kettenbruchprobleme, *Compositio Math.*, **1** (1935), 361—382.
2. Zur metrischen Kettenbruchtheorie, *Compositio Math.*, **3** (1936), 275—285.
3. Об основных теоремах теории информации, *УМН*, **11 : 1**, **№ 67** (1956), 17—75.
4. Цепные дроби, Физматгиз, М., 1961.

Хопф Э.
1. Эргодическая теория, *УМН*, **4**, вып. 1 (1949), 113—182.

Чжун Кай-лай (Chung K. L.)
1. A note on the ergodic theorem of information theory, *Ann. Math. Statist.*, **32** (1961), 612—614.

Шеннон К. Э.
1. Математическая теория связи, 1948, стр. 243—332.

Эгглстон (Eggleston H. G.)
1. The fractional dimension of a set defined by decimal properties, *Quart. J. Math. Oxford Ser.*, **20** (1949), 31—36.

Якобс (Jacobs K.)
1. Neuere Methoden und Ergebnisse der Ergodentheorie, Springer, 1960.
2. Lecture notes on ergodic theory, University of Aarhus, 1962—1963.

Указатель примеров

Пример 1.1 (стр. 11). *Двусторонний сдвиг Бернулли.*

Ω: $\omega = (\ldots, \omega_{-1}, \omega_0, \omega_1, \ldots)$, где $x_n(\omega) = \omega_n$ — элемент конечного множества ρ,

$T\omega = (\ldots, \omega_0, \omega_1, \omega_2, \ldots)$, т. е. $x_n(T\omega) = x_{n+1}(\omega)$,

$$P\{\omega\colon x_l(\omega) = i_l,\ n \leqslant l \leqslant n+k\} = \prod_{l=n}^{n+k-1} p_{i_l}.$$

Свойства: обратимый (стр. 13); эргодический (стр. 22); перемешивающий (стр. 22); $h(T) = -\sum_i p_i \ln p_i$ (стр. 80 и 102).

Пример 1.2 (стр. 13). *Общий двусторонний сдвиг.*

Ω: $\omega = (\ldots, \omega_{-1}, \omega_0, \omega_1, \ldots)$, где $x_n(\omega) = \omega_n$ — элемент конечного множества ρ,

$T\omega = (\ldots, \omega_0, \omega_1, \omega_2, \ldots)$, т. е. $x_n(T\omega) = x_{n+1}(\omega)$,

P — любая вероятностная мера, сохраняющаяся при преобразовании T.

Свойства: обратимый.

Пример 1.3 (стр. 14). *Нециклическая перестановка.*

$\Omega = \{a, b, c, d, e\}$,

$T = (a, b, c)(d, e)$,

$P(a) = P(b) = P(c)$ и $P(d) = P(e)$.

Свойства: преобразование обратимое; неэргодическое, если только $P(a) = P(b) = P(c)$ не равны 0 или $P(d) = P(e)$ не равны 0 (стр. 16); не перемешивающее (стр. 22); $h(T) = 0$.

Пример 1.4 (стр. 14). *Циклическая перестановка.*

$\Omega = \{a, b, c, d, e\}$,

$T = (a, b, c, d, e)$,

$P(a) = P(b) = P(c) = P(d) = P(e) = \frac{1}{5}.$

Свойства: обратимое; эргодическое (стр. 16); не перемешивающее (стр. 22); $h(T)=0$.

Пример 1.5 (стр. 15). *Поворот окружности.*

Ω — единичная окружность в комплексной плоскости,
$T\omega = c\omega$, где c — элемент из Ω,
P — нормированная круговая мера Лебега.

Свойства: обратимое; эргодическое в том и только том случае, если c не является корнем из единицы (стр. 18); не перемешивающее (стр. 22); $h(T)=0$ (стр. 103 и 104).

Пример 1.6 (стр. 15). *Диадическое преобразование.*

$\Omega=[0, 1]$
$T\omega = 2\omega \pmod 1$,
P — мера Лебега.

Свойства: необратимое; эргодическое (стр. 20 и 119); перемешивающее (стр. 22); $h(T)=\ln 2$.

Пример 1.7 (стр. 28). *Сдвиг для вещественнозначного процесса.*

Ω: $\omega = (\ldots, \omega_{-1}, \omega_0, \omega_1, \ldots)$, где $x_n(\omega)=\omega_n$ — точка вещественной прямой,
$T\omega = (\ldots, \omega_0, \omega_1, \omega_2, \ldots)$, т. е. $x_n(T\omega)=x_{n+1}(\omega)$,
P — любая вероятностная мера, сохраняющаяся при преобразовании T.

Свойства: обратимое.

Пример 3.1 (стр. 41). *Сдвиг Маркова.*

Частный случай общего двустороннего сдвига (пример 1.2), задаваемого формулой

$$P\{\omega\colon x_{n+l-1}(\omega)=i_l,\ l=1, \ldots, k\} = p_{i_1} p_{i_1 i_2} \ldots p_{i_{k-1} i_k},$$

где p_i для матрицы вероятностей перехода (p_{ij}) стационарны и строго положительны.

Свойства: обратимый; эргодический в том и только том случае, если матрица (p_{ij}) неприводима (стр. 42, 43, 131 и 138); перемешивающий в том и только том случае, если матрица (p_{ij}) неприводима и непериодична (стр. 44 и 140); $h(T) = -\sum_{ij} p_i p_{ij} \ln p_{ij}$ (стр. 102).

Пример 3.2 (стр. 44). *Общий односторонний сдвиг.*

Ω: $\omega = (\omega_1, \omega_2, \ldots)$, где $x_n(\omega) = \omega_n$ — элемент конечного множества ρ,
$T\omega = (\omega_2, \omega_3, \ldots)$, т. е. $x_n(T\omega) = x_{n+1}(\omega)$,
P — любая вероятностная мера, сохраняющаяся при преобразовании T.

Свойства: никогда не обратим; свойства эргодичности и перемешивания, а также значения энтропии совпадают с аналогичными свойствами и значениями для соответствующего двустороннего сдвига.

Пример 3.3 (стр. 45).

Частный случай общего двустороннего сдвига (пример 1.2), в котором исходы независимы, но каждый повторяется.

Свойства: обратимый; эргодический (стр. 45); не перемешивающий (стр. 45); $h(T) = -\frac{1}{2}\sum_i p_i \ln p_i$ (стр. 103).

Пример 3.4 (стр. 45). *r-адическое преобразование.*

$\Omega = [0, 1)$,
$T\omega = r\omega \pmod 1$,
P — мера Лебега.

Свойства: необратимое; эргодическое, перемешивающее; $h(T) = \ln r$.

Пример 3.5 (стр. 46).

$\Omega = [0, 1)$,
$T\omega = r\omega \pmod 1$,
P — любая вероятностная мера, сохраняющаяся при преобразовании T.

Свойства: необратимо.

ДОПОЛНЕНИЕ

Алгебраические автоморфизмы тора и цепи Маркова

Б. М. Гуревич и Я. Г. Синай

Как отмечено в предисловии редактора перевода, важность изложенной в книге теории объясняется не только ее связями с теорией вероятностей, но и возможностью успешного применения к изучению динамических систем, порождаемых диффеоморфизмами и векторными полями на гладких многообразиях. В этом дополнении мы подробно рассматриваем один весьма популярный пример такого рода — алгебраический автоморфизм тора — и устанавливаем, что с точки зрения эргодической теории он изоморфен марковскому автоморфизму (см. стр. 41) с конечным или счетным пространством состояний. Читателю предоставляется возможность самому судить, какая из областей — алгебра или теория вероятностей — от этого больше выигрывает.

§ 1. ПРЕДВАРИТЕЛЬНЫЕ СВЕДЕНИЯ ОБ АВТОМОРФИЗМЕ n-МЕРНОГО ТОРА

Тор M получается из n-мерного евклидова пространства R^n, рассматриваемого как топологическая группа, факторизацией по целочисленной решетке Z^n. Геометрически тор можно представлять себе как единичный n-мерный куб $\{0 \leqslant y_i \leqslant 1;\ i = 1, 2, \ldots, n\}$ с мерой Лебега μ, у которого отождествлены противоположные $(n-1)$-мерные грани. Расстояние $d(x_1, x_2)$ между точками $x_1, x_2 \in M$ определяется равенством

$$d(x_1, x_2) = \rho(C_{x_1}, C_{x_2}), \tag{1}$$

где ρ — метрика в R^n, а C_{x_1}, C_{x_2} — классы смежности R^n по Z^n, отвечающие точкам x_1, x_2.

Любая целочисленная матрица A порядка n переводит решетку Z^n в себя. Если, кроме того, $|\det A| = 1$, то A^{-1} обладает тем же свойством и, следовательно, A отображает Z^n на себя взаимно однозначно. В таком случае матрица A индуцирует групповой автоморфизм тора $M = R^n/Z^n$, сохраняющий меру μ. Всякий групповой автоморфизм тора получается описанным здесь способом.

Будем обозначать π естественную проекцию (естественный гомоморфизм) пространства R^n на тор, т. е. отображение, сопоставляющее каждой точке $y \in R^n$ содержащий ее класс смежности по подгруппе Z^n. Если A_T — матрица, индуцирующая автоморфизм T, то $\pi A_T y = T\pi y$, $y \in R^n$. Формула (1) с помощью π может быть записана в виде

$$d(x_1, x_2) = \min_{y_1, y_2 : \pi y_1 = x_1, \pi y_2 = x_2} \rho(y_1, y_2).$$

Пусть $\lambda_1, \lambda_2, \ldots, \lambda_s$ — собственные значения матрицы A_T и $n_1, n_2, \ldots, n_s$ — их кратности. Так как матрица вещественна, то каждому вещественному λ_i отвечает комплексно-сопряженное число $\lambda_j = \bar{\lambda}_i$, имеющее ту же кратность, что и λ_i. Приведя матрицу A_T к жордановой форме, мы получим для каждого вещественного λ_i n-мерное инвариантное подпространство H_i пространства R^n и в H_i базис, в котором A_T имеет вид жордановой клетки $\begin{pmatrix} \lambda_i & & & \\ 1 & \lambda_i & & 0 \\ & 1 & \ddots & \\ & & \ddots & \\ 0 & & 1 & \lambda_i \end{pmatrix}$. Если $\lambda_j, \bar{\lambda}_j$ — пара комплексно-сопряженных собственных значений, то можно найти $2n$-мерное инвариантное подпространство H_j и в нем базис, в котором матрица имеет вид жордановой клетки

$$\begin{pmatrix} \Lambda_j & & & & \\ E & \Lambda_j & & & 0 \\ & E & \ddots & & \\ & & \ddots & \ddots & \\ & & & \ddots & \\ 0 & & & E & \Lambda_j \end{pmatrix}, \quad \text{где} \quad \Lambda_j = \begin{pmatrix} \operatorname{Re}\lambda_j & \operatorname{Im}\lambda_j \\ -\operatorname{Im}\lambda_j & \operatorname{Re}\lambda_j \end{pmatrix}, \quad E = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Назовем *расширяющимся подпространством* и обозначим H_p прямую сумму подпространств H_i по тем i, для которых $|\lambda_i| > 1$. Аналогично *сжимающееся подпространство* H_c есть прямая сумма тех H_i, для которых $|\lambda_i| < 1$. Из определения видно, что H_p и H_c инвариантны относительно A_T. В дальнейшем *мы будем рассматривать лишь такие матрицы A_T, которые не имеют собственных значений, равных по модулю единице.* Это условие, очевидно, эквивалентно равенству

$$H_p \oplus H_c = R^n.$$

Так как $|\det A_T|=1$, то оба подпространства $H_\text{р}$ и $H_\text{с}$ в рассматриваемом случае нетривиальны.

Пусть $\lambda = \prod\limits_{i:|\lambda_i|>1} |\lambda_i|$ и $\mu_\text{р}$, $\mu_\text{с}$ — лебеговы меры соответственно на $H_\text{р}$ и $H_\text{с}$. Пользуясь каноническим видом матрицы A_T, заключаем, что для любых измеримых множеств $B_\text{р} \subset H_\text{р}$ и $B_\text{с} \subset H_\text{с}$

$$\mu_\text{р}(A_T B_\text{р}) = \lambda\mu_\text{р}(B_\text{р}), \quad \mu_\text{с}(A_T B_\text{с}) = \lambda^{-1}\mu_\text{с}(B_\text{с}). \tag{2}$$

Выясним теперь поведение длин векторов в пространствах $H_\text{р}$ и $H_\text{с}$ под действием степеней матрицы A_T. Длину вектора $y \in R^n$ будем обозначать $|y|$.

Лемма 1[1]). *Существуют такие константы* $a_\text{с}$, $a_\text{р}$, $\lambda_\text{с}<1$, $\lambda_\text{р}>1$, *что для любых векторов* $y'' \in H_\text{с}$, $y' \in H_\text{р}$ *и любого натурального* m

$$|A_T^m y''| \leqslant a_\text{с}\lambda_\text{с}^m |y''|, \quad |A_T^{-m} y'| \leqslant a_\text{р}\lambda_\text{р}^{-m} |y'|. \tag{3}$$

Доказательство. Так как подпространства $H_\text{с}$ и $H_\text{р}$ меняются ролями при замене A_T на A_T^{-1}, достаточно доказать первое из неравенств (3). Воспользуемся каноническим представлением матрицы A_T в подпространстве $H_\text{с}$. Возьмем любое $q \neq 0$ и сопоставим клеткам

$$\begin{bmatrix} \lambda_i & & & & 0 \\ 1 & \lambda_i & & & \\ & 1 & \ddots & & \\ & & \ddots & \ddots & \\ 0 & & & 1 & \lambda_i \end{bmatrix} \quad \text{и} \quad \begin{bmatrix} \Lambda_j & & & & 0 \\ E & \Lambda_j & & & \\ & E & \ddots & & \\ & & \ddots & \ddots & \\ 0 & & & E & \Lambda_j \end{bmatrix}$$

в этом представлении соответственно клетки

$$\begin{bmatrix} 1 & & & & 0 \\ & q & & & \\ & & q^2 & & \\ & & & \ddots & \\ 0 & & & & q^{n_i-1} \end{bmatrix} \quad \text{и} \quad \begin{bmatrix} E & & & & 0 \\ & Q & & & \\ & & Q^2 & & \\ & & & \ddots & \\ 0 & & & & Q^{n_j-1} \end{bmatrix},$$

---

[1]) Из этой леммы вытекает, что если матрица A_T не имеет собственных значений, по модулю равных единице, то автоморфизм T является $У$-системой в смысле Аносова (см. [2]).

где $Q = qE$. Матрица, состоящая из этих клеток, приводит A_T к виду, получающемуся из канонического заменой элементов 1 и E соответственно на q и Q. Пусть $e_1, e_2, \ldots, e_k$ — базис, в котором A_T имеет только что описанный вид. Введем в H_c новую норму, положив $\| y'' \|^2 = \sum_{i=1}^{k} c_i^2$, если $y'' = \sum_{i=1}^{n} c_i e_i$. Пользуясь тем, что все собственные значения, относящиеся к H_c, по модулю меньше единицы, нетрудно проверить, что если q достаточно мало, то $\| A_T \| < 1$ (в подпространстве H_c). Другими словами, можно так подобрать $\lambda_c < 1$, чтобы для любого $y'' \in H_c$

$$\| A_T y'' \| < \lambda_c \| y'' \|$$

и, следовательно,

$$\| A_T^m y'' \| < \lambda_c^m \| y'' \|, \qquad m = 0, 1, \ldots.$$

Интересующее нас неравенство (3) вытекает теперь из эквивалентности любых двух норм в конечномерном линейном пространстве.

Рассмотрим в пространстве R^n гиперплоскости, параллельные подпространству H_p, и назовем их *расширяющимися*. Преобразование A_T переводит расширяющиеся гиперплоскости друг в друга. То же самое можно сказать и о *сжимающихся* гиперплоскостях, параллельных подпространству H_c. Образы при отображении π гиперплоскостей каждого семейства будем называть соответственно *расширяющимися* и *сжимающимися слоями* (на торе). На любой расширяющейся или сжимающейся гиперплоскости отображение π взаимно однозначно. Достаточно проверить это для расширяющихся гиперплоскостей, так как сжимающиеся гиперплоскости являются расширяющимися по отношению к A_T^{-1}. Очевидно, можно ограничиться подпространством H_p. Пусть y_1', y_2' — векторы из H_p и $\pi y_1' = \pi y_2'$. Это значит, что $y_3' = y_1' - y_2'$ — целочисленный вектор, также принадлежащий H_p. По лемме 1 $|A_T^m y_3'| \to 0$ при $m \to -\infty$, что невозможно, если $y_3' \neq 0$ (в этом случае A_T^m — ненулевой целочисленный вектор). Тем самым наше утверждение доказано.

Л е м м а 2. *Любой расширяющийся или сжимающийся слой всюду плотен на торе.*

Д о к а з а т е л ь с т в о. Здесь снова достаточно рассмотреть слой πH_p. Пусть $\overline{\pi H_p}$ — его замыкание. Множество $\overline{\pi H_p}$ является замкнутой связной подгруппой тора M, инвариантной относительно

сительно автоморфизма T. Факторгруппа $M/\pi\overline{H_p}$ имеет тогда, как известно, вид $M/\pi H$, где H — замкнутая подгруппа группы R^n, содержащая H_p и целочисленную решетку Z^n и инвариантная относительно A_T. При этом обе группы $M/\pi H$ и R^n/H изоморфны тору некоторой размерности $k<n$ (см. [4, гл. 7, § 1, п. 5]). Покажем, что $k=0$.

Введем метрику ρ' в группе R^n/H, положив расстояние между двумя ее элементами равным расстоянию в R^n между соответствующими смежными классами. Пользуясь инвариантностью подгруппы H относительно A_T, перенесем это преобразование в R^n/H, где оно останется гомеоморфизмом. Возьмем любые два смежных класса C_1 и C_2 из R^n/H. Нетрудно найти точки $y_1 \in C_1$ и $y_2 \in C_2$, проекции которых на H_p совпадают. Отсюда следует (см. лемму 1), что

$$\rho'\left(A_T^m C_1,\ A_T^m C_2\right) \to 0 \quad \text{при} \quad m \to \infty.$$

Но в силу компактности группы R^n/H это возможно лишь в случае, когда множество R^n/H состоит из одной точки, что и требовалось доказать.

§ 2. МАРКОВСКИЕ РАЗБИЕНИЯ И ЭРГОДИЧЕСКИЕ СВОЙСТВА АВТОМОРФИЗМОВ ТОРА

В этом параграфе мы введем марковские разбиения и докажем с их помощью ряд эргодических свойств автоморфизмов n-мерного тора. Существование марковских разбиений будет установлено в следующем параграфе. Напомним еще раз, что по предположению все собственные значения рассматриваемой матрицы A_T не равны по модулю единице.

Пусть V', V'' — измеримые ограниченные множества в подпространствах H_p и H_c соответственно, $\operatorname{Int} V'$, $\operatorname{Int} V''$ — их открытые ядра (по отношению к H_p и H_c) и $\partial V' = \overline{V'} \setminus \operatorname{Int} V'$, $\partial V'' = \overline{V''} \setminus \operatorname{Int} V''$ — их границы (черта обозначает замыкание). Назовем множество

$$V = (V',\ V'') = \{y \in R^n\colon\ y = y' + y'',\ y' \in V',\ y'' \in V''\}\,^{1)}$$

параллелограммом в R^n, если $\mu_p(\partial V') = \mu_c(\partial V'') = 0$. Образ параллелограмма $V \subset R^n$ при отображении π будем также называть параллелограммом, если π на множестве V является

---

[1]) Здесь y, y', y'' обозначают n-мерные векторы. В дальнейшем будет употребляться запись $y = (y',\ y'')$.

гомеоморфизмом. В этом случае

$$\mu(\pi V)=\gamma\mu_p(V')\mu_c(V''), \tag{4}$$

где γ — константа, зависящая от угла между подпространствами H_p и H_c. Легко видеть, что пересечение двух параллелограммов в R^n есть снова параллелограмм.

В дальнейшем мы будем пользоваться следующей леммой, доказательство которой несложно и предоставляется читателю.

Лемма 3. *Пусть* $\delta=\max(|A_T|, |A_T^{-1}|)$ *и* U_1, U_2 — *связные параллелограммы на торе, диаметры которых меньше* $1/4\delta$, *причем* $U_2=\pi V_2$, *где* V_2 — *связный параллелограмм в* R^n. *Тогда найдется такой связный параллелограмм* $V_1\subset R^n$, *что* $U_1=\pi V_1$, $U_1\cap TU_2=\pi(V_1\cap A_TV_2)$ *и на множестве* $V_1\cup A_TV_2$ *отображение* π *является изометрией и, следовательно, гомеоморфизмом. Утверждение останется справедливым, если заменить в нем* T *и* A_T *соответственно на* T^{-1} *и* A_T^{-1}.

Граница ∂V *параллелограмма* $V\subset R^n$ совпадает с объединением множеств $\Gamma_p(V)=(\bar{V}', \partial V'')$ и $\Gamma_c(V)=(\partial V', \bar{V}'')$. Первое из этих множеств будем называть *расширяющейся границей*, второе — *сжимающейся*. Аналогично множества $\pi\Gamma_p(V)=\Gamma_p(\pi V)$ и $\pi\Gamma_c(V)=\Gamma_c(\pi V)$ образуют соответственно расширяющуюся и сжимающуюся границы параллелограмма πV.

Будем называть *расширяющимся слоем параллелограмма* $V\subset R^n$ пересечение любой расширяющейся гиперплоскости с V и *граничным расширяющимся слоем* — пересечение расширяющейся гиперплоскости с $\Gamma_p(V)$. Аналогично определим *сжимающийся слой* и *граничный сжимающийся слой параллелограмма* V. Отображение π переводит все указанные слои в одноименные слои параллелограмма $U=\pi V$.

Определим теперь *разбиение тора на параллелограммы*. Так мы будем называть конечную или бесконечную последовательность α параллелограммов $U_i\subset M$, удовлетворяющую трем условиям: 1) $\overline{\bigcup_i U_i}=M$, 2) $\operatorname{Int} U_i\cap\operatorname{Int} U_j=\varnothing$ при $i\neq j$, 3) $\mu\left(\bigcup_i \operatorname{Int} U_i\right)=1$. Заметим, что α не является, вообще говоря, разбиением в строгом смысле слова. В частности, его элементы могут пересекаться. Однако все их пересечения содержатся в множестве $\Gamma(\alpha)=\overline{\bigcup_i \partial U_i}$, имеющем нулевую меру. В случае когда α конечно, условие 1) можно записать в виде

$\bigcup\limits_i \overline{U}_i = M$, а необходимость в условии 3) отпадает, так как оно вытекает из 1) и определения параллелограмма. Положим

$$\Gamma_p(\alpha) = \overline{\bigcup \Gamma_p(U_i)}, \quad \Gamma_c(\alpha) = \overline{\bigcup_i \Gamma_c(U_i)}.$$

В случае конечного α очевидно, что

$$\Gamma_p(\alpha) = \bigcup_i \Gamma_p(U_i), \quad \Gamma_c(\alpha) = \bigcup_i \Gamma_c(U_i).$$

Определение (см. [5]). Разбиение α на параллелограммы называется *марковским*, если $\Gamma_p(T^{-1}\alpha) \subset \Gamma_p(\alpha)$, $\Gamma_c(T\alpha) \subset \Gamma_c(\alpha)$.

В следующем параграфе мы покажем, что рассматриваемый *автоморфизм T обладает конечным марковским разбиением с элементами произвольно малого диаметра.* Отсюда вытекает утверждение, на котором будет основано дальнейшее изучение автоморфизмов тора.

Теорема 1. *Для всякого $\varepsilon > 0$ существует марковское (по отношению к T) разбиение β, элементы которого B_i связны, открыты и удовлетворяют условию* $\operatorname{diam} B_i \leqslant \varepsilon$, $i = 1, 2, \ldots$.

Доказательство. Пусть U — произвольный элемент конечного марковского разбиения α. Открытое ядро $\operatorname{Int} U$ параллелограмма U также является параллелограммом. Разобьем множество $\operatorname{Int} U$ на связные компоненты, их может быть конечное или счетное число. Нетрудно проверить, что каждое из получившихся множеств — открытый параллелограмм и эти множества образуют разбиение тора на параллелограммы. Обозначим это разбиение β. Тот факт, что оно марковское, вытекает из марковского свойства разбиения α и соотношений $\Gamma_p(\beta) = \Gamma_p(\alpha)$, $\Gamma_c(\beta) = \Gamma_c(\alpha)$.

Основной результат этого параграфа содержит

Теорема 2. 1) *Автоморфизм T изоморфен сдвигу Маркова с конечным или счетным множеством состояний.* 2) $h(T) = \ln \lambda$ $\left(\text{\textit{напомним, что}}\ \lambda = \prod\limits_{i:\,|\lambda_i| > 1} \lambda_i\right)$. 3) *$T$ является K-автоморфизмом* (см. определение на стр. 111).

Доказательство. 1) Пусть β — марковское разбиение, описанное в теореме 1, $B_1, B_2, \ldots$ — его элементы, являющиеся открытыми связными параллелограммами, причем

$\operatorname{diam} B_i < 1/2\delta$, $i = 1, 2, \ldots$ (δ определено в лемме 3). Введем матрицу пересечений $S = (s_{ij})$, где

$$s_{ij} = \begin{cases} 1, & \text{если} \quad B_i \cap TB_j \neq \varnothing, \\ 0, & \text{если} \quad B_i \cap TB_j = \varnothing, \end{cases}$$

и рассмотрим в пространстве Ω бесконечных в обе стороны последовательностей натуральных чисел множество Ω_S, состоящее из тех последовательностей $\omega = \{\omega_k\}$, для которых $s_{\omega_k, \omega_{k+1}} = 1$, $-\infty < k < \infty$. Нетрудно проверить, что Ω_S измеримо, т. е. принадлежит σ-алгебре $\mathscr{B}$, порожденной цилиндрическими множествами, и инвариантно относительно сдвига $\widetilde{T}$, переводящего последовательность $\omega = \{\omega_k\}$ в $\omega' = \{\omega'_k\}$, где $\omega'_k = \omega_{k-1}$.

Покажем, что для любого $m \geqslant 0$ и любой последовательности $\omega = \{\omega_k\} \in \Omega_S$ в R^n найдутся такие параллелограммы $C_{\omega_0} = (C'_{\omega_0}, C''_{\omega_0})$, $C_{\omega_m} = (C'_{\omega_m}, C''_{\omega_m})$, что

$$B_{\omega_0} = \pi C_{\omega_0}, \quad B_{\omega_m} = \pi C_{\omega_m}, \tag{5}$$

$$\bigcap_{k=0}^{m} T^k B_{\omega_k} = \pi \left(C'_{\omega_0}, A_T^m C''_{\omega_m}\right), \quad A_T^m C''_{\omega_m} \subset C''_{\omega_0}.$$

Пусть $m = 1$. Пользуясь леммой 3, найдем в R^n параллелограммы $C_{\omega_0} = (C'_{\omega_0}, C''_{\omega_0})$ и $C_{\omega_1} = (C'_{\omega_1}, C''_{\omega_1})$, для которых

$$B_{\omega_0} = \pi C_{\omega_0}, \quad B_{\omega_1} = \pi C_{\omega_1}, \quad B_{\omega_0} \cap TB_{\omega_1} = \pi (C_{\omega_0} \cap A_T C_{\omega_1})$$

и на $C_{\omega_0} \cup A_T C_{\omega_1}$ отображение π является гомеоморфизмом. Тогда множества C'_{ω_0}, C''_{ω_0}, $A_T C'_{\omega_1}$, $A_T C''_{\omega_1}$ связны и открыты (в соответствующих подпространствах) и $C'_{\omega_0} \cap A_T C'_{\omega_1} \neq \varnothing$, $C''_{\omega_0} \cap A_T C''_{\omega_1} \neq \varnothing$. В силу марковского свойства разбиения β

$$\Gamma_c(TB_{\omega_1}) \subset \Gamma_c(\beta),$$

а потому

$$\partial (A_T C'_{\omega_1}) \cap C'_{\omega_0} = \varnothing. \tag{6}$$

Отсюда следует, что

$$C'_{\omega_0} \subset A_T C'_{\omega_1}. \tag{7}$$

В самом деле, если это не так, возьмем любую точку из $C'_{\omega_0} \cap (H_p \setminus A_T C'_{\omega_1})$ и соединим ее непрерывной кривой, лежащей в C'_{ω_0}, с какой-нибудь точкой из $C'_{\omega_0} \cap A_T C'_{\omega_1}$. На этой кривой обязана найтись точка из $A_T(\partial C'_{\omega_1}) = \partial (A_T C'_{\omega_1})$, что противоречит равенству (6).

Аналогичные рассуждения, но с использованием условия $\Gamma_p(T^{-1}B_{\omega_0}) \subset \Gamma_p(\beta)$, показывают, что

$$A_T C''_{\omega_1} \subset C''_{\omega_0}. \tag{8}$$

Из (7) и (8) вытекает, что

$$C_{\omega_0} \cap A_T C_{\omega_1} = (C'_{\omega_0} \cap A_T C'_{\omega_1},\ C''_{\omega_0} \cap A_T C''_{\omega_1}) = (C'_{\omega_0},\ A_T C''_{\omega_1}), \tag{9}$$

т. е. при $m=1$ наше утверждение справедливо. По индукции легко доказать его для любого $m>0$, пользуясь инвариантностью множества Ω_S относительно T.

Совершенно так же устанавливается, что в R^n существуют параллелограммы $D_{\omega_0} = (D'_{\omega_0}, D''_{\omega_0})$ и $D_{\omega_{-m}} = (D'_{\omega_{-m}}, D''_{\omega_{-m}})$, для которых

$$\begin{gathered} B_{\omega_0} = \pi D_{\omega_0}, \quad B_{\omega_{-m}} = \pi D_{\omega_{-m}}, \\ \bigcap_{k=0}^{-m} T^k B_{\omega_k} = \pi(A_T^{-m} D'_{\omega_{-m}},\ D''_{\omega_0}), \quad A_T^{-m} D'_{\omega_{-m}} \subset D_{\omega_0}. \end{gathered} \tag{10}$$

При этом можно считать, что $D_{\omega_0} = C_{\omega_0}$. Заменив обозначение $D_{\omega_{-m}}$ на $C_{\omega_{-m}}$, комбинируя (5) и (10), получим равенство $\bigcap_{k=-m}^{m} T^k B_{\omega_k} = \pi(A_T^{-m} C'_{\omega_{-m}},\ A_T^m C''_{\omega_m})$, из которого следует, что множество $B_m(\omega) = \bigcap_{k=-m}^{m} T^k B_{\omega_k}$ непусто и $\operatorname{diam} B_m(\omega) \leqslant (a_c \lambda_c^m + a_p \lambda_p^{-m})/2\delta$ (см. леммы 1 и 3). Тогда $\bigcap_{|m|<\infty} B_m(\omega)$ — единственная точка тора M, и мы можем определить отображение φ множества Ω_S в M, положив

$$\varphi(\omega) = \bigcap_{|m|<\infty} \overline{B_m(\omega)}, \qquad \omega \in \Omega_S.$$

Это отображение удовлетворяет условию

$$\varphi(\tilde{T}\omega) = T\varphi(\omega), \qquad \omega \in \Omega_S. \tag{11}$$

Покажем, что оно измеримо. Пусть $\bar{B}_m$ — замкнутое множество вида $\bigcap_{|i| \leqslant m} T^i B_{k_i}$, где B_{k_i}, $-m \leqslant i \leqslant m$, — произвольные элементы разбиения β. Если $\bar{B}_m \neq \varnothing$, то в Ω_S найдется точка ω, для которой $\overline{B_m(\omega)} = \bar{B}_m$. Поэтому $\operatorname{diam} \bar{B}_m = \operatorname{diam} B_m(\omega) \leqslant \varepsilon_m$, где $\varepsilon_m = (a_c \lambda_c^m + a_p \lambda_p^{-m})/2\delta$. Занумеруем

при фиксированном m множества $\overline{B}_m$ в последовательность $\overline{B}_m^1, \overline{B}_m^2, \ldots$ и положим $C_m^1 = \overline{B}_m^1$, $C_m^2 = \overline{B}_m^2 \setminus C_m^1, \ldots, C_m^k = \overline{B}_m^k \setminus \bigcup_{i=1}^{k-1} C_m^k, \ldots$. Множества C_m^k, $k = 1, 2, \ldots$, не пересекаются и образуют разбиение множества $\bigcup_{k=1}^{\infty} \overline{B}_m^k$ на конечное или счетное число борелевских подмножеств, диаметры которых не превосходят ε_m. Так как $\mu\left(\bigcup_{k=1}^{\infty} \overline{B}_m^k\right) = 1$, то множество $M \setminus \bigcup_{k=1}^{\infty} \overline{B}_m^k$ можно разбить на счетное число борелевских подмножеств произвольно малого диаметра. В итоге мы получаем разбиение ξ_m тора M на счетное число борелевских множеств, диаметры которых не превосходят ε_m, причем $\lim_{n\to\infty} \varepsilon_m = 0$. Из определения φ видно, что при любых m и k множество $\varphi^{-1}(\overline{B}_m^k)$ есть пересечение некоторого цилиндра с Ω_S, т. е. $\varphi^{-1}(\overline{B}_m^k) \in \tilde{\mathscr{B}}$. Отсюда следует, что прообраз любого элемента разбиения ξ_m принадлежит $\tilde{\mathscr{B}}$, и для доказательства измеримости φ достаточно установить, что совокупность элементов всех ξ_m, $m = 1, 2, \ldots$, порождает σ-алгебру $\mathscr{B}$ борелевских подмножеств тора. Но это непосредственно вытекает из доказанной в книге леммы 3 § 5[1]).

Итак, отображение φ измеримо, но, как легко понять, не взаимно однозначно. Однако для φ можно найти обратное отображение, если предварительно выкинуть из тора некоторое множество меры 0. В самом деле, объединение $\bigcup_{k=1}^{\infty} B_k$ элементов разбиения β есть открытое множество меры 1. Поэтому множество $D = \bigcap_{|i| < \infty} T^i \bigcup_{k=1}^{\infty} B_k$, инвариантное относительно T, является борелевским и также имеет меру 1.

---

[1]) Если ввести σ-алгебру $\mathscr{B}_0$, атомами которой служат параллелограммы $B_1, B_2, \ldots$ и множество $M \setminus \bigcup_{k=1}^{\infty} B_k$, то доказанное утверждение можно сформулировать в виде равенства $\bigvee_k \bigvee_{i=-k}^{k} T^i \mathscr{B}_0 \stackrel{*}{=} \mathscr{B}$ (определение символа $\stackrel{*}{=}$ см. на стр. 104).

В силу измеримости φ отсюда следует, что $\tilde{D} = \varphi^{-1}(D) \in \tilde{\mathscr{B}}$. Заметим, что любая точка $\omega \in D$ имеет единственный прообраз $\varphi^{-1}(\omega) \in \Omega_S$. Поэтому на D можно определить отображение ψ, обратное к φ и переводящее D в $\tilde{D}$. Это отображение измеримо, так как для любого цилиндра $\tilde{F}$ множество $\psi^{-1}(\tilde{F} \cap \tilde{D}) = \varphi(\tilde{F} \cap \tilde{D}) = \varphi(\tilde{F}) \cap \varphi(\tilde{D}) = \varphi(\tilde{F}) \cap D$ является борелевским.

Зададим на σ-алгебре $\tilde{\mathscr{B}}$ меру $\tilde{\mu}$ формулой

$$\tilde{\mu}(\tilde{B}) = \tilde{\mu}(\tilde{B} \cap \tilde{D}) = \mu(\varphi(\tilde{B} \cap \tilde{D})), \quad \tilde{B} \in \tilde{\mathscr{B}}.$$

Мера $\tilde{\mu}$ инвариантна относительно $\tilde{T}$ и сосредоточена на инвариантном множестве $\tilde{D}$. Вместе с равенством (11) это означает, что преобразования $(T, \mathscr{B}, \mu)$ и $(\tilde{T}, \tilde{\mathscr{B}}, \tilde{\mu})$ изоморфны (см. определение на стр. 66).

Покажем, что $\tilde{T}$ — сдвиг Маркова. Пусть $B_i = \pi C_i$, $C_i = (C_i', C_i'')$, $\mu_p(C_i') = b_i'$, $\mu_c(C_i'') = b_i''$ и $p_i = \mu(B_i)$, $p_{ij} = b_j'' s_{ij} / \lambda b_i''$. Из (5) с учетом (3) и (4) вытекает, что если $B_i \cap TB_j \neq \varnothing$, то

$$\mu(B_i \cap TB_j) = \frac{\gamma}{\lambda} \mu_c(C_j'') \mu_p(C_i') = \mu(B_i) \frac{\mu_c(C_j'')}{\lambda \mu_c(B_i'')} = \mu(B_i) \frac{b_j''}{\lambda b_i''},$$

т. е. для цилиндра $\tilde{F} = \{\omega \colon \omega_0 = i,\ \omega_1 = j\}$

$$\tilde{\mu}(\tilde{F}) = \mu(\varphi(\tilde{F} \cap \tilde{D})) = \mu(\varphi(\tilde{F}) \cap \varphi(\tilde{D})) = \\ = \mu(\varphi(\tilde{F}) \cap D) = \mu(\varphi(\tilde{F})) = \mu(B_i \cap TB_j) = p_i p_{ij}.$$

Отсюда, суммируя по j, заключаем, что матрица $P = (p_{ij})$ стохастическая.

Чтобы доказать нужное нам утверждение, достаточно проверить для любых элементов $B_{i_0}, B_{i_1}, \ldots, B_{i_m}$ разбиения β равенство

$$\mu\left(\bigcap_{k=0}^{m} T^k B_{i_k}\right) = p_{i_0} \prod_{k=0}^{m-1} p_{i_k i_{k+1}}, \tag{12}$$

которое, как мы уже видели, справедливо при $m = 1$. Если $s_{i_k i_{k+1}} = 0$ хотя бы при одном $k < m$, то обе части равенства (12) обращаются в 0. Если же $s_{i_k i_{k+1}} = 1$ при $k = 0, 1, \ldots, m-1$, то правая часть равна $\gamma b_{i_0}' b_{i_m}'' \lambda^{-m}$, а левая в силу (5) есть $\mu\left(\pi\left(C_{\omega_0}', A_T^m C_{\omega_m}''\right)\right) = \gamma b_{i_0}' b_{i_m}'' \lambda^{-m}$. Тем самым равенство (12) доказано.

2) Чтобы вычислить энтропию $h(T)$, вернемся к конечному марковскому разбиению α. Если его элементы U_1, $U_2, \ldots, U_s$ достаточно малы, то всякое пересечение вида $\bigcap_{i=j}^{k} T^{-i} U_{l_i}$ $(j=0,\ 1;\ k=1,\ 2\ \ldots)$ представляет собой параллелограмм. Пусть x — произвольная точка инвариантного множества B и $U(x,\ i)$ — элемент разбиения α, содержащий $T^i x$. Рассмотрим параллелограмм $C_j^k(x) = \bigcup_{i=j}^{k} T^{-i} U(x,\ i)$. По лемме 1 при $k \to \infty$ диаметр любого расширяющегося слоя этого параллелограмма стремится к нулю, а потому множество $C_j(x) = \bigcap_k C_j^k(x)$ содержится в сжимающемся слое параллелограмма $T^{-j} U(x,\ j)$, проходящем через точку x. С другой стороны, из соотношения между марковскими разбиениями α и β вытекает, что $C_j(x)$ содержит проходящий через x сжимающийся слой того элемента разбиения $T^{-j}\beta$, в котором лежит x.

С помощью проекции π, которая, как мы знаем, взаимно однозначна на каждой сжимающейся гиперплоскости, перенесем меру Лебега с этих гиперплоскостей на соответствующие сжимающиеся слои, сохранив для получившейся меры на слое обозначение μ_c. Из построения множеств $C_j(x)$ видно, что каждое такое множество измеримо на содержащем его сжимающемся слое и функции $\mu_j(x) = \mu_c(C_j(x))$, $j = 0,\ 1$, определенные на множестве полной меры, измеримы.

Рассмотрим алгебру $\mathcal{A}$ подмножеств тора, атомами которой служат параллелограммы $U_1, U_2, \ldots, U_s$ и множество $M \setminus \bigcup_{i=1}^{s} U_i$. При всяком $l = 1,\ 2, \ldots, s$ условная мера $\mu\left(U_l \,\middle|\, \bigvee_{i=1}^{\infty} T^{-i}\mathcal{A}\right)$ почти всюду равна $\mu_c(U_l \cap C_1(x))/\mu_1(x)$. В самом деле, согласно теореме 11.2, достаточно проверить, что почти всюду

$$\lim_{k\to\infty} \mu\left(U_l \,\middle|\, \bigvee_{i=1}^{k} T^{-i}\mathcal{A}\right) = \frac{\mu_c(U_l \cap C_1(x))}{\mu_1(x)}. \tag{13}$$

В любой точке $x \in B$ условная мера $\mu\left(U_l \,\middle|\, \bigvee_{i=1}^{k} T^{-i}\mathcal{A}\right)$ совпадает с $\mu(U_l \mid C_1^k(x))$. Если при всех k, начиная с некоторого, множества U_l и $C_1^k(x)$ не пересекаются, то обе части равен-

ства (13) равны нулю. Предположим, что $U_l \cap C_1^k(x) \neq \varnothing$ при бесконечно многих k. Так как $x \in B$, то x — внутренняя точка параллелограмма U_l. Поэтому при достаточно большом k любой расширяющийся слой параллелограмма $C_1^k(x)$, пересекающийся с U_l, содержится в U_l. Отсюда следует (см. (4)), что

$$\mu\left(U_l \mid C_1^k(x)\right) = \frac{\mu_c\left(U_l \cap L_1^k(x)\right)}{\mu_c\left(L_1^k(x)\right)},$$

где $L_1^k(x)$ — сжимающийся слой параллелограмма $C_1^k(x)$, проходящий через точку x, и так как $C_1(x) = \bigcap\limits_k L_1^k(x)$, то функции $\mu_c\left(L_1^k(x)\right)$ и $\mu_c\left(U_l \cap L_1^k(x)\right)$ сходятся при $k \to \infty$ соответственно к $\mu_c(C_1(x)) = \mu_1(x)$ и $\mu_c(U_l \cap C_1(x))$. Тем самым равенство

$$\mu\left(U_l \,\Big|\, \bigvee_{i=1}^{\infty} T^{-i}\mathcal{A}\right) = \frac{\mu_c(U_l \cap C_1(x))}{\mu_1(x)}$$

доказано.

Теперь заметим, что по построению $C_0(x) = TC_1\left(T^{-1}x\right)$ при $x \in B$ и $U_l \cap C_1(x) = C_0(x)$ при $x \in B \cap U_l$. Поэтому при $x \in B \cap U_l$

$$\mu\left(U_l \,\Big|\, \bigvee_{i=1}^{\infty} T^{-i}\mathcal{A}\right) = \frac{\mu_0(x)}{\mu_1(x)} = \frac{\mu_1(T^{-1}x)}{\lambda\mu_1(x)}.$$

Отсюда, пользуясь соотношением (12.2) (см. стр. 141), получим

$$h(T, \mathcal{A}) = h\left(\mathcal{A} \,\Big|\, \bigvee_{i=1}^{\infty} T^{-i}\mathcal{A}\right) = -\int\limits_M \ln \frac{\mu_1(T^{-1}x)}{\lambda\mu_1(x)}\, d\mu = \ln\lambda + {}$$
$$+ \int\limits_M \left(\ln\mu_1(x) - \ln\mu_1\left(T^{-1}x\right)\right) d\mu.$$

Левая часть этого равенства конечна, так как $h(T, \mathcal{A}) \leqslant H(\mathcal{A})$. Значит, функция $g(x) = f(x) - f\left(T^{-1}x\right)$, где $f(x) = \ln\mu_1(x)$, интегрируема. По теореме Биркгофа левая часть равенства

$$\frac{1}{N}\sum_{i=0}^{N-1} g(T^i x) = \frac{1}{N}\left(f\left(T^{N-1}x\right) - f\left(T^{-1}x\right)\right)$$

при $N \to \infty$ сходится почти всюду к функции $\hat{g}(x)$, для которой $\int\limits_M \hat{g}(x)\,d\mu = \int\limits_M g(x)\,d\mu$. Правая же часть, как нетрудно понять, может сходиться только к нулю. Следовательно, $\int\limits_M g(x)\,d\mu = 0$, $h(T, \mathcal{A}) = \ln\lambda$.

В заключение установим, что

$$h(T, \mathcal{A}) = h(T).$$

Рассмотрим конечную алгебру $\mathcal{A}_k = \bigvee\limits_{i=-k}^{k} T^i\mathcal{A}$, $k = 1, 2, \ldots$. Любой атом этой алгебры, имеющий положительную меру, очевидно, представим в виде $\bigcup\limits_{i=-k}^{k} T^i U_{l_i}$. Пусть δ_k — максимальный из диаметров таких атомов. Если параллелограммы U_1, $U_2, \ldots, U_s$ достаточно малы, то, повторяя рассуждения предыдущего пункта, можно показать, что $\lim\limits_{k\to\infty}\delta_k = 0$ и $\bigvee\limits_{k=1}^{\infty} \mathcal{A}_k \overset{*}{=} \mathscr{B}$. Поэтому (см. следствие 1 на стр. 105) $h(T) = \lim\limits_{k\to\infty} h(T, \mathcal{A}_k)$, и для завершения доказательства остается заметить, что $h(T, \mathcal{A}_k) = h(T, \mathcal{A})$ при любом k. В итоге получаем

$$h(T) = h(T, \mathcal{A}) = \ln\lambda.$$

3) Для того чтобы доказать, что T является K-автоморфизмом, рассмотрим вместо T сдвиг $\tilde{T}$. Если марковское разбиение β конечно, то, согласно § 11 книги, достаточно проверить неприводимость и апериодичность матрицы P. То же верно и при бесконечном β.

В самом деле, для неприводимой апериодической матрицы P выполняются соотношения $\lim\limits_{k\to\infty} p_{ij}^{(k)} = p_j$ $(i, j = 1, 2, \ldots)$, где $p_{ij}^{(k)}$ — элементы матрицы P^k (см. [6, стр. 384—385]). Отсюда, пользуясь свойствами цепи Маркова и теоремой Дуба (см. стр. 138), нетрудно вывести, что если $\tilde{\mathscr{B}}_k$ есть σ-алгебра событий, зависящих от поведения рассматриваемой цепи Маркова в моменты $l \geqslant k$, а $\tilde{F}$ — произвольный цилиндр, то почти всюду $\lim\limits_{k\to\infty} \tilde{\mu}(\tilde{F} \mid \tilde{\mathscr{B}}_k) = \tilde{\mu}(\tilde{F})$.

Пусть $\tilde{B}$ — любое множество из σ-алгебры $\tilde{\mathscr{B}}_\infty = \bigcap\limits_k \tilde{\mathscr{B}}_k$ и $I_{\tilde{B}}$ — его характеристическая функция. Для $\tilde{B}$ (как и для

всякого измеримого множества) можно подобрать такую последовательность множеств $\tilde{B}_k$, что каждое $\tilde{B}_k$ есть объединение конечного числа непересекающихся цилиндров и почти всюду $I_{\tilde{B}_k} \to I_{\tilde{B}}$ при $k \to \infty$, т. е. почти всюду

$$\lim_{k \to \infty} E\left(I_{\tilde{B}_k} \middle| \tilde{\mathscr{B}}_\infty\right) = E\left(I_{\tilde{B}} \middle| \tilde{\mathscr{B}}_\infty\right) = I_{\tilde{B}}.$$

По доказанному $E\left(I_{\tilde{B}_k} \middle| \tilde{\mathscr{B}}_\infty\right) = \tilde{\mu}(\tilde{B}_k)$. Это означает, что почти всюду $I_{\tilde{B}} = \mu(\tilde{B})$ и, следовательно, $\tilde{\mu}(\tilde{B}) = 0$ или 1. Таким образом, T является K-автоморфизмом, если матрица P неприводима и апериодична.

Заметим, что оба эти условия заведомо выполняются, если для любых B_i, B_j найдется такой номер $k_0(i, j)$, что $B_i \cap T^k B_j \neq \varnothing$ при всех $k \geqslant k_0(i, j)$. Существование для нашей матрицы чисел $k_0(i, j)$ с указанным свойством мы установим, опираясь на лемму 2. Из нее вытекает, что для всякого $\varepsilon > 0$ найдется такое число R_ε, что любой шар радиуса R_ε, лежащий в расширяющемся слое[1]), является ε-сетью для тора. В самом деле, пусть L — какой-нибудь расширяющийся слой. Так как по доказанному он всюду плотен на торе, шары радиуса $\varepsilon/2$ с центрами в точках слоя L образуют покрытие тора. Выделив конечное подпокрытие, мы получим конечную $(\varepsilon/2)$-сеть, состоящую из центров выделенных шаров. Шар $O'(x)$, лежащий в слое L, с центром в заданной точке $x \in L$ и достаточно большого радиуса $R(x, \varepsilon/2)$, содержит эту $(\varepsilon/2)$-сеть, а потому и сам является $(\varepsilon/2)$-сетью. Для любой точки $x \in M$ обозначим $L(x)$ расширяющийся слой, проходящий через точку x. Пусть x_1, $x_2, \ldots, x_l$ — произвольная конечная $(\varepsilon/2)$-сеть и O' — шар радиуса $R = R_\varepsilon = \max\limits_{1 \leqslant k \leqslant l} R(x_k, \varepsilon/2)$ с центром в произвольной точке x, лежащий в слое $L(x)$. Найдем точку x_k, для которой $d(x, x_k) \leqslant \varepsilon/2$. Шар $O'(x_k)$, лежащий в слое $L(x_k)$, образует $(\varepsilon/2)$-сеть. Из неравенства $R \geqslant R(x_k, \varepsilon/2)$ следует тогда, что шар O' образует ε-сеть. Наше утверждение доказано.

Пусть теперь B_i, B_j — любые элементы разбиения β. Каждое из этих множеств, будучи открытым, содержит некоторый шар. Обозначим эти шары соответственно O_i и O_j, а их центры — x_i и x_j. Пересечение шара O_j с расширяю-

---

[1]) Так естественно называть множество вида πO, где O — пересечение n-мерного шара радиуса R_ε с расширяющейся гиперплоскостью, содержащей центр этого шара.

щимся слоем $L(x_j)$ есть некоторый шар O'_j, лежащий в слое $L(x_j)$. Пусть $\varepsilon = r/2$, где r — радиус шара O_i. Найдем соответствующее R_ε. На основании леммы 1 существует такое k_0, что при $k > k_0$ множество $T^k O'_j$ содержит шар радиуса R_ε, лежащий в слое $L(T^k x_j)$. По доказанному этот шар является ε-сетью и, значит, пересекается с шаром O_i. Но $O_i \subset B_i$ и $T^k O'_j \subset T^k B_j$. Следовательно, при $k > k_0$ пересечение $B_i \cap T^k B_j$ непусто. Теорема полностью доказана.

§ 3. СУЩЕСТВОВАНИЕ МАРКОВСКОГО РАЗБИЕНИЯ

1. Марковские разбиения для T и T^m. Мы назвали разбиение α на параллелограммы *марковским* для автоморфизма T, если оно удовлетворяет условиям $\Gamma_p(T^{-1}\alpha) \subset \Gamma_p(\alpha)$, $\Gamma_c(T\alpha) \subset \Gamma_c(\alpha)$. Прежде чем строить такое разбиение, заметим, что достаточно найти марковское разбиение α' для T^m при каком-нибудь $m > 0$[1]). В самом деле, пусть

$$\Gamma_p(T^{-m}\alpha') \subset \Gamma_p(\alpha'), \quad \Gamma_c(T^m\alpha') \subset \Gamma_c(\alpha'). \tag{14}$$

Множества вида

$$T^{-m+1}U_{i_{-m+1}} \cap T^{-m+2}U_{i_{-m+2}} \cap \ldots \cap T^{-1}U_{i_{-1}} \cap U_{i_0} \cap TU_{i_1} \cap \ldots \cap T^{m-1}U_{i_{m-1}},$$

где U_{i_k} — произвольные элементы разбиения α', очевидно, также образуют разбиение тора на параллелограммы. Для этого разбиения введем обозначение $\alpha = T^{-m+1}\alpha' \vee \ldots \vee T^{-1}\alpha' \vee \vee \alpha' \vee T\alpha' \vee \ldots \vee T^{m-1}\alpha'$. Легко проверяются следующие соотношения между границами α и α':

$$\Gamma_p(\alpha) = \bigcup_{k=-m+1}^{m-1} \Gamma_p(T^k\alpha'), \quad \Gamma_c(\alpha) = \bigcup_{k=-m+1}^{m-1} \Gamma_c(T^k\alpha'). \tag{15}$$

Из (14) и (15) вытекает, что α — марковское разбиение для T, причем размеры его элементов не превышают размеров соответствующих элементов разбиения α'.

2. Разбиение, полученное из покрытия. Построим какое-нибудь конечное *покрытие тора параллелограммами*, т. е.

[1]) Подпространства H_p и H_c, а вместе с ними параллелограммы и границы параллелограммов, очевидно, от m не зависят.

систему u параллелограммов $U_1, U_2, \ldots, U_s$, дающих в сумме весь тор. Такая система получится, если включить каждую точку тора в какой-нибудь открытый параллелограмм, а затем выделить в случае необходимости конечное подпокрытие. Данный способ позволяет получить покрытие, размеры элементов которого произвольно малы.

Опишем теперь один общий способ перехода от покрытия к разбиению, который в дальнейшем будет неоднократно применяться. Пусть $\tau = \{X_i,\ i = 1, 2, \ldots, s\}$ — произвольное конечное семейство подмножеств некоторого множества X и ξ_i — разбиение этого множества, состоящее из двух элементов: X_i и $X \setminus X_i$. Сопоставим семейству τ разбиение $\xi(\tau)$, где $\xi(\tau) = \bigvee_{i=1}^{s} \xi_i$ [1]. Взяв в качестве τ совокупность u параллелограммов U_i, $i = 1, 2, \ldots, s$, образующих покрытие тора, мы получим разбиение $\xi(u)$. Элементы этого разбиения уже не обязаны быть параллелограммами.

Построим, исходя из покрытия u, разбиение тора на параллелограммы. Пусть U_i — произвольный элемент нашего покрытия и $U_{i_1}, U_{i_2}, \ldots, U_{i_k}$ — элементы покрытия, пересекающиеся с U_i. Если размеры элементов покрытия u достаточно малы, то в пространстве R^n найдутся такие параллелограммы $V_i = (V'_i, V''_i)$; $V_{i_1} = (V'_{i_1}, V''_{i_1}), \ldots, V_{i_k} = (V'_{i_k}, V''_{i_k})$, что $U_i = \pi V_i$, $U_{i_1} = \pi V_{i_1}, \ldots, U_{i_k} = \pi V_{i_k}$ и отображение π на множестве $V_i \cup V_{i_1} \cup \ldots \cup V_{i_k}$ является гомеоморфизмом. Отсюда, в частности, следует, что каждое множество из семейств $v'_i = \{V'_{i_l} \cap V'_i,\ l = 1, 2, \ldots, k\}$ и $v''_i = \{V''_{i_l} \cap V''_i,\ l = 1, 2, \ldots, k\}$ непусто. Построим описанным уже способом разбиение $\xi(v'_i)$ множества V'_i и разбиение $\xi(v''_i)$ множества V''_i. Пусть C' — произвольный элемент первого из этих разбиений, а C'' — второго. Легко видеть, что $\mu_p(\partial C') = \mu_c(\partial C'') = 0$, и мы получаем некоторое разбиение ζ_i параллелограмма V_i на параллелограммы вида $C = (C', C'')$. Разбиение параллелограмма $U_i = \pi V_i$, образованное параллелограммами вида πC, обозначим η_i.

---

[1]) Если $\xi_1, \xi_2, \ldots, \xi_s$ — разбиения множества X, то произведение $\bigvee_{i=1}^{s} \xi_i$ есть по определению разбиение множества X, образованное множествами вида $\bigcap_{i=1}^{s} C_{k_i}$, где C_{k_i} — произвольный элемент разбиения ξ_i.

Любой элемент разбиения $\xi(u)$, содержащийся в параллелограмме U_i, состоит из элементов разбиения η_i. Другими словами, каждый элемент разбиения η_i содержится в некотором элементе разбиения $\xi(u)$. Действительно, пусть точки $x_1, x_2 \in U_i$ принадлежат разным элементам разбиения $\xi(u)$ и $x_1 = \pi y_1$, $x_2 = \pi y_2$, $y_1, y_2 \in V_i$, $y_1 = (y_1', y_1'')$, $y_2 = (y_2', y_2'')$. Это значит, что среди параллелограммов $V_{i_1}, V_{i_2}, \ldots, V_{i_k}$ найдется такой параллелограмм V_j, что либо $y_1 \in V_j$, $y_2 \notin V_j$, либо $y_1 \notin V_j$, $y_2 \in V_j$. В каждом из случаев либо точки y_1', y_2' принадлежат разным элементам разбиения $\xi(v_i')$, либо точки y_1'', y_2'' — разным элементам разбиения $\xi(v_i'')$. Но тогда точки y_1, y_2 принадлежат разным элементам разбиения ζ_i и, следовательно, точки x_1, x_2 — разным элементам разбиения η_i.

Пользуясь доказанным утверждением, рассмотрим на каждом элементе $C_{\xi(u)}$ разбиения $\xi(u)$ произведение разбиений η_i по тем i, для которых $C_{\xi(u)} \subset U_i$. Полученное таким образом разбиение тора на параллелограммы обозначим $\alpha(u)$. Заметим, что $\alpha(u)$ является разбиением в строгом смысле слова, но некоторые из его элементов могут не иметь внутренних точек. Выбросив эти элементы и взяв открытые ядра остальных, мы можем при желании получить разбиение тора на открытые параллелограммы в смысле § 2. Это разбиение также будем обозначать $\alpha(u)$.

3. Границы разбиения $\alpha(u)$. Пусть $\Gamma_p(u) = \bigcup_{i=1}^{s} \Gamma_p(u_i)$, $\Gamma_c(u) = \bigcup_{i=1}^{s} \Gamma_c(u_i)$ — соответственно *расширяющаяся и сжимающаяся границы покрытия u*. Сейчас мы установим связь между $\Gamma_p(u)$, $\Gamma_c(u)$ и соответствующими границами $\Gamma_p(\alpha(u))$, $\Gamma_c(\alpha(u))$ разбиения $\alpha(u)$.

Предположим, что элементы покрытия u и элементы разбиения $\alpha(u)$ — открытые множества. Тогда связь между $\Gamma_p(u)$ и $\Gamma_p(\alpha(u))$ выражается следующим образом: $\Gamma_p(\alpha(u))$ *состоит из всех граничных расширяющихся слоев элементов покрытия u и тех расширяющихся слоев, которые пересекаются с граничными расширяющимися слоями.* Аналогичное утверждение относится к сжимающейся границе. Очевидно, доказательство достаточно провести лишь для $\Gamma_p(\alpha(u))$.

Включение $\Gamma_p(u) \subset \Gamma_p(\alpha(u))$ следует из того, что по построению $\Gamma_p(U_i) \subset \Gamma_p(\eta_i) \subset \Gamma_p(\alpha(u))$ для любого i. Покажем, что любой расширяющийся слой параллелограмма U_i, пере-

секающийся с граничным расширяющимся слоем некоторого параллелограмма U_j, также входит в $\Gamma_p(\alpha(u))$. Вместо U_i, U_j будем рассматривать параллелограммы V_i, V_j. Пусть $L_i = (V_i', y_i'')$, $L_j = (\bar{V}_j', y_j'')$ — соответствующие расширяющиеся слои. Тогда $y_i'' = y_j''$ и $V_i' \cap V_j' \neq \varnothing$. Из построения видно, что точка $y'' = y_i'' = y_j''$ принадлежит границе одного из элементов разбиения $\xi(v_i'')$, а множество V_i' состоит из элементов разбиения $\xi(v_i')$. Поэтому всякая точка слоя $L_i = (V_i', y'')$ принадлежит $\Gamma_p(\zeta_i)$ и, следовательно, $\pi L_i \subset \Gamma_p(\eta_i) \subset \Gamma_p(\alpha(u))$.

Обратно, пусть $x \in \Gamma_p(\alpha(u)) \setminus \Gamma_p(u)$. Тогда $x \in U_i \cap \Gamma_p(\eta_i)$ при некотором i, и если $y = (y', y'') = \pi^{-1}x$ — соответствующая точка параллелограмма V_i, то $y \in \Gamma_p(\zeta_i)$. Отсюда вытекает, что точка y'' принадлежит границе некоторого элемента разбиения $\xi(v_i'')$. Значит, среди параллелограммов $V_{i_1}, V_{i_2}, \ldots, V_{i_k}$, пересекающихся с V_i, найдется такой параллелограмм $V_j = (V_j', V_j'')$, что $y'' \in \partial V_j''$. Но тогда расширяющийся слой $L_i = (V_i', y'')$, содержащий точку y, пересекается с граничным расширяющимся слоем $L_j = (\bar{V}_j', y'')$. Тем самым наше утверждение доказано.

4. Марковское разбиение и марковское покрытие. По аналогии с марковским разбиением введем понятие *марковского покрытия*. Покрытие u тора параллелограммами мы будем называть *марковским*, если

$$\Gamma_p(T^{-1}u) \subset \Gamma_p(u), \quad \Gamma_c(Tu) \subset \Gamma_c(u).$$

В предыдущих пунктах описан метод, позволяющий перейти от покрытия параллелограммами к разбиению на параллелограммы. Выясним теперь, при каких условиях этот метод приводит к марковскому разбиению.

Рассмотрим покрытие $u = (U_1, U_2, \ldots, U_s)$ тора M параллелограммами достаточно малого диаметра. Пусть $x \in U_i$ и $L_p(x, i)$, $L_c(x, i)$ — соответственно расширяющийся и сжимающийся слои параллелограмма U_i, содержащие точку x. Назовем покрытие u *правильным* относительно T, если при некотором $\varepsilon > 0$ оно удовлетворяет двум условиям:

1) для любой точки $x \in M$ найдется содержащий ее параллелограмм U_i, такой, что

$$d(x, \partial L_p(x, i)) > \varepsilon, \quad d(x, \partial L_c(x, i)) > \varepsilon;$$

2) для любого $i = 1, 2, \ldots, s$ и любой точки $x \in U_i$

$$\operatorname{diam}(T^{-1}L_p(x, i)) < \varepsilon, \quad \operatorname{diam}(TL_c(x, i)) < \varepsilon.$$

Покажем, что *если u — правильное марковское покрытие, то $\alpha(u)$ — марковское разбиение.*

Здесь, как и раньше, достаточно проверить лишь одно из требуемых включений, например $\Gamma_c(T\alpha(u)) \subset \Gamma_c(\alpha(u))$ или, что то же самое, $T\Gamma_c(\alpha(u)) \subset \Gamma_c(\alpha(u))$. Если $x \in \Gamma_c(u)$, то включение $Tx \in \Gamma_c(\alpha(u))$ очевидно, так как $\Gamma_c(u) \subset \Gamma_c(\alpha(u))$ и покрытие u — марковское. Пусть $x \in \Gamma_c(\alpha(u)) \setminus \Gamma_c(u)$. Пользуясь результатом предыдущего пункта, найдем параллелограмм U_i и в нем сжимающийся слой $L_i = L_c(x, i)$, содержащий x и пересекающийся с граничным сжимающимся слоем L_j некоторого другого параллелограмма U_j. В силу правильности покрытия u для точки Tx найдется такой параллелограмм U_k и в нем сжимающийся слой L_k, что $Tx \in L_k$ и $d(Tx, \partial L_k) > \varepsilon$. Отсюда видно, что $TL_i \subset TL_k$ (так как $\operatorname{diam}(TL_i) < \varepsilon$) и, значит, $TL_j \cap L_k \neq \varnothing$. Но в силу марковского свойства покрытия u любая точка множества TL_j принадлежит некоторому граничному сжимающемуся слою. Это означает, что L_k пересекается с таким слоем. Следовательно, слой L_k, а с ним и точка Tx, содержится в $\Gamma_c(\alpha(u))$.

Таким образом, мы свели вопрос о существовании марковского разбиения для автоморфизма T к вопросу о существовании правильного марковского покрытия. Однако условия 1) и 2), определяющие правильное покрытие, до некоторой степени противоположны друг другу, и это затрудняет построение правильного марковского покрытия непосредственно для T.

Выход подсказывается результатом пункта 1, согласно которому достаточно получить марковское разбиение для автоморфизма T^m при каком-нибудь $m > 0$, что в свою очередь сводится к построению для T^m правильного марковского покрытия.

5. Построение правильного марковского покрытия для T^m. Пусть x — любая точка тора и $y = (y', y'')$ — один из ее прообразов при отображении π. Рассмотрим при достаточно малом $\varepsilon > 0$ параллелограмм $O_\varepsilon(y) = (O'_\varepsilon, O''_\varepsilon) \subset R^n$, где O'_ε, O''_ε — открытые шары радиуса ε в подпространствах H_p, H_c с центрами в точках y', y'' соответственно. Параллелограммы вида $\pi O_\varepsilon(y)$ образуют открытое покрытие тора. Выделим из него конечное подпокрытие $\pi O_\varepsilon(y_1)$, $\pi O_\varepsilon(y_2)$, ..., $\pi O_\varepsilon(y_s)$ и после этого заменим каждый параллелограмм $O_\varepsilon(y_i)$ параллелограммом $V_i = O_{2\varepsilon}(y_i)$, $i = 1, 2, \ldots, s$. Тогда параллелограммы $U_i = \pi V_i$, $i = 1, 2, \ldots, s$, также образуют покрытие тора, которое удовлетворяет первому из требований, предъ-

являемых к правильному покрытию. Теперь можно, пользуясь леммой 1, подобрать настолько большое m, что покрытие $u=(U_1, U_2, \ldots, U_s)$ станет правильным по отношению к автоморфизму T^m (выбор m мы в дальнейшем еще уточним).

Остается превратить наше покрытие в марковское для T^m, не нарушив его правильности. Для начала заметим, что в проведенном построении точки $y_i=(y_i', y_i'')$ можно было выбрать так, чтобы выполнялось условие

$$\rho(0, y_i'')<\frac{1}{4}\varepsilon; \tag{16}$$

это непосредственно вытекает из леммы 2. Каждый параллелограмм V_i, а с ним и $A_T^mV_i$, пересекается тогда с подпространством H_ρ. Рассмотрим вместе с каждым V_i, $i=1, 2, \ldots$ $\ldots, s$, все его сдвиги на целочисленные векторы. Получившиеся s семейств параллелограммов образуют в совокупности некоторое покрытие v пространства R^n, причем любое ограниченное множество в R^n пересекается лишь с конечным числом элементов этого покрытия. Покрытие v обладает, очевидно, теми же свойствами правильности по отношению к A_T^m, что и u (по отношению к T^m). Поэтому для каждого параллелограмма $A_T^mV_i=(A_T^mV_i', A_T^mV_i'')$ можно найти такие элементы $V_{ij}=(V_{ij}', V_{ij}'')$, $j=1, 2, \ldots, s_i$, покрытия v, что

$$A_T^mV_i'\subset\bigcup_{j=1}^{s_i}V_{ij}', \quad \rho(0, \partial V_{ij}')>\varepsilon \tag{17}$$

и, следовательно, $A_T^mV_i''\subset V_{ij}''$ при каждом $j=1, 2, \ldots, s_i$ в силу правильности v. По условию параллелограмм V_{ij} получается из некоторого V_l, $1\leqslant l\leqslant s$, сдвигом на подходящий целочисленный вектор $z_l=(z_l', z_l'')$, т. е. V_{ij} имеет вид V_l+z_l. Сравнивая (16) и (17), получаем оценку

$$|z_l''|<\frac{5}{4}\varepsilon. \tag{18}$$

Заменим теперь обозначение V_i' на $V_i'(0)$ и V_{ij}' на $V_{ij}'(0)$. Пусть $v'(0)$ — совокупность множеств $V_i'(0)$, $w'(0)$ — совокупность множеств $V_{ij}'(0)$, $i=1, 2, \ldots, s$, $j=1, 2, \ldots, s_i$, и $\Gamma(v'(0))=\bigcup_i\partial V_i'(0)$, $\Gamma(w'(0))=\bigcup_{i, j}\partial V_{ij}'(0)$. Будем строить индуктивно последовательности $\{v'(k)\}$ и $\{w'(k)\}$, $k=0, 1, \ldots,$

где $v'(k)$ — совокупность множеств $V'_i(k)$, $i=1, 2, \ldots, s$, $w'(k)$ — совокупность множеств $V'_{ij}(k)$, $j=1, 2, \ldots, s_i$, и $V'_{ij}(k)=V'_i(k)+z'_i$. Если $v'(k)$ и $w'(k)$ уже построены, то положим

$$V'_i(k+1)=A_T^{-m}\left(\bigcup_{j=1}^{s_i} V'_{ij}(k)\right). \tag{19}$$

Из этого определения вытекает, что для любых i, j последовательности множеств $\{V'_i(k)\}$ и $\{V'_{ij}(k)\}$ возрастают. Пусть $V'_i(\infty)=\bigcup\limits_k V'_i(k)$, $V'_{ij}(\infty)=\bigcup\limits_k V'_{ij}(k)$, $v'(\infty)$ — совокупность множеств $V'_i(\infty)$ и $w'(\infty)$ — совокупность множеств $V'_{ij}(\infty)$.

Займемся изучением свойств построенных множеств. Для этого введем величины

$$d'(k)=\max_i \operatorname{diam} V'_i(k), \quad k=0, 1, \ldots, \infty,$$

$$r'(k)=\max_i \sup_{y'\in\partial V'_i(k+1)} \rho(y', \partial V'_i(k)), \quad k=0, 1, \ldots,$$

которые, очевидно, связаны между собой неравенством $d'(k+1)\leqslant d'(k)+2r'(k)$. По построению $d'(0)=4\varepsilon$. Из леммы 1 и соотношения (19) нетрудно вывести, что

$$r'(0)\leqslant 4a_p\lambda_p^{-m}\varepsilon,$$

$$r'(k+1)\leqslant a_p\lambda_p^{-m}r'(k), \quad k=0, 1, \ldots,$$

и, следовательно,

$$\begin{gathered} r'(k)\leqslant 4\varepsilon\theta_p^{k+1}, \\ d'(k)\leqslant 4\varepsilon\left(1+2\theta_p\sum_{l=0}^{k-1}\theta_p^l\right), \end{gathered} \tag{20}$$

где $\theta_p=a_p\lambda_p^{-m}$. Выбором достаточно большого m можно добиться, чтобы было $\theta_p<1/7$. Тогда

$$d'(\infty)=\lim_{k\to\infty} d'(k)=4\varepsilon\left(1+2\theta_p\sum_{l=0}^{\infty}\theta_p^l\right)=4\varepsilon\left(1+\frac{2\theta_p}{1-\theta_p}\right)<6\varepsilon$$

и для любого $i=1, 2, \ldots, s$

$$\operatorname{diam}\left(A_T^{-m}V'_i(\infty)\right)\leqslant\theta_p\operatorname{diam}\left(V'_i(\infty)\right)<\varepsilon. \tag{21}$$

Теперь покажем, что

$$A_T^m\Gamma(v'(\infty))\subset\Gamma(w'(\infty)) \tag{22}$$

и

$$\mu_\rho(\Gamma(v'(\infty)))=0. \tag{23}$$

Включение (22) следует из соотношений

$$A_T^m V_i'(\infty)=\bigcup_{k=1}^{\infty} A_T^m V_i'(k)=\bigcup_{k=1}^{\infty}\bigcup_{j=1}^{s_i} V_{ij}'(k-1)=$$

$$=\bigcup_{j=1}^{s_i}\bigcup_{k=1}^{\infty} V_{ij}'(k-1)=\bigcup_{j=1}^{s_i} V_{ij}'(\infty),$$

$$A_T^m\partial\left(V_i'(\infty)\right)=\partial\left(A_T^m V_i'(\infty)\right)=$$

$$=\partial\left(\bigcup_{j=1}^{s_i} V_{ij}'(\infty)\right)\subset\bigcup_{j=1}^{s_i}\partial\left(V_{ij}'(\infty)\right)\subset\Gamma(w'(w)).$$

Равенство (23) доказывается сложнее. Прежде всего заметим, что множество $A_T^{mk}V_i'(k)$ можно представить в виде

$$\bigcup_j\left(V_{l_j}'(0)+z_{l_j,0}'+A_T^m z_{l_j,1}'+\ldots+A_T^{(k-1)m}z_{l_j,k-1}'\right), \tag{24}$$

где $z_{l_j,0}', z_{l_j,1}', \ldots, z_{l_j,k-1}'$ — проекции на H_ρ целочисленных векторов $z_{l_j,0}, z_{l_j,1}, \ldots, z_{l_j,k-1}$. Это легко проверить по индукции, вспомнив способ построения $V_i'(k)$ (см. (19)). Равенство (24), очевидно, останется справедливым, если в левой его части заменить $V_i'(k)$ на $V_i'(k+1)$, а в правой $V_{l_j}'(0)$ на $V_{l_j}'(1)$. Отсюда и из (20) вытекает, что для любой точки $y'\in\partial A_T^{km}V_i'(k+1)$

$$\rho\left(y',\partial A_T^{km}V_i'(k)\right)\leqslant r'(0)\leqslant 4\theta_\rho\varepsilon.$$

Пусть $r(k,l)=\sup\limits_{y'\in A_T^{km}V_i'(k+l)}\rho\left(y',\partial A_T^{km}V_i'(k)\right)$. Тогда при всех $k,\ l>0$

$$r(k,l)\leqslant r(k,1)+\sup_{y'\in A_T^{km}V_i'(k+l)}\rho\left(y',\partial A_T^{km}V_i'(k+1)\right)\leqslant$$

$$\leqslant r(k,1)+\theta_\rho\sup_{y'\in A_T^{(k+1)m}V_i'(k+l)}\rho\left(y',\partial A_T^{(k+1)m}V_i'(k+1)\right)=$$

$$=r(k,1)+\theta_\rho r(k+1,l-1)=\ldots$$

$$\ldots=r(k,1)+\theta_\rho r(k+1,1)+\ldots+\theta_\rho^{l-1}r(k+l-1,1)\leqslant$$

$$\leqslant 4\theta_\rho\varepsilon\left(1+\theta_\rho+\ldots+\theta_\rho^{l-1}\right)<\frac{4\theta_\rho\varepsilon}{1-\theta_\rho}<\frac{2}{3}\varepsilon.$$

Переходя к пределу при $l \to \infty$ и фиксированном k, получаем

$$r(k, \infty) = \sup_{y' \in \partial A_T^{km} V_i'(\infty)} \rho(y', \partial A_T^{km} V_i'(k)) \leqslant \frac{2}{3}\varepsilon. \qquad (25)$$

Вернемся к покрытию v пространства R^n. Пересечения элементов этого покрытия с H_p образуют некоторое покрытие v' подпространства H_p. Построим разбиение $\xi(v')$ (см. п. 2). Из (16), (18) и (24) вытекает, что если m достаточно велико, то при любых i, k множество $A_T^k V_i'(k)$ состоит из элементов $\xi(v')$. Действительно, рассмотрим вместе с множеством $V' = V_{l_j}'(0) + z_{l_j,0}' + A_T^m z_{l_j,1}' + \ldots + A_T^{(k-1)m} z_{l_j,k-1}'$ параллелограмм $V = V_{l_j}(0) + z_{l_j,0} + A_T^m z_{l_j,1} + \ldots + A_T^{(k-1)m} z_{l_j,k-1}$. В силу (18)

$$\left| z_{l_j,0}'' + A_T^m z_{l_j,1}'' + \ldots + A_T^{(k-1)m} z_{l_j,k-1}'' \right| \leqslant$$
$$\leqslant \frac{5}{4}\varepsilon\left(1 + \theta_c + \theta_c^2 + \ldots + \theta_c^{k-1}\right) \leqslant \frac{5}{4}\varepsilon\frac{1}{1-\theta_c}, \qquad (26)$$

где $\theta_c = a_c \lambda_c^m$ (см. лемму 1). Если m выбрано так, что $\theta_c < 1/7$, то в силу (16) и (26) $V \cap H_p \neq \varnothing$ и, следовательно, $V \cap H_p = V'$. Остается вспомнить, что по определению разбиения $\xi(v')$ множество $V \cap H_p$ состоит из элементов этого разбиения.

Установим теперь еще одно необходимое нам свойство $\xi(v')$: существует конечное число элементов разбиения $\xi(v')$, сдвигая которые на векторы из H_p, можно получить все остальные элементы этого разбиения.

Сначала заметим, что аналогичным свойством обладает покрытие v'. Именно, для любого его элемента V' найдется такое i, $1 \leqslant i \leqslant s$, что $V' = V_i'(0) + z_i'$, где $V_i'(0)$ — шар радиуса 2ε, принадлежащий системе $v'(0)$, и z_i' — вектор из H_p. Зафиксировав для каждого V' так подобранное i и выделив элементы с одним и тем же i, разобьем совокупность элементов нашего покрытия на s классов, причем элементы одного класса будут получаться друг из друга с помощью сдвигов на векторы из H_p.

Оценим снизу длины этих векторов и их разностей. Пусть V_i', $\tilde{V}_i'$ и V_j', $\tilde{V}_j'$ — элементы соответственно i-го и j-го классов. Тогда найдутся такие параллелограммы V_i, $\tilde{V}_i$, V_j, $\tilde{V}_j$, являющиеся элементами покрытия v, что $V_i' = V_i \cap H_p$, $\tilde{V}_i' = \tilde{V}_i \cap H_p$, $V_j' = V_j \cap H_p$, $\tilde{V}_j' = \tilde{V}_j \cap H_p$, $\tilde{V}_i = V_i + z_i$, $\tilde{V}_j = V_j + z_j$, где $z_i = (z_i', z_i'')$ и $z_j = (z_j', z_j'')$ — целочисленные векторы. Пользуясь тем, что все рассматриваемые параллелограммы пересекаются

с H_p, а их проекции на H_c имеют диаметр 4ε, легко проверить, что если векторы z'_i, z'_j, $z'_i - z'_j$ ненулевые, то

$$|z'_i| \geqslant 1 - 8\varepsilon, \quad |z'_j| \geqslant 1 - 8\varepsilon, \quad |z'_i - z'_j| \geqslant 1 - 16\varepsilon. \tag{27}$$

Отсюда, в частности, следует, что при $\varepsilon < 1/12$ элементы одного класса не могут пересекаться друг с другом.

Рассмотрим произвольный элемент C разбиения $\xi(v')$. По определению C есть пересечение некоторых элементов покрытия v' с дополнениями к остальным элементам покрытия. При $\varepsilon < 1/16$ элементы покрытия v', содержащие C, принадлежат разным классам. Значит, таких элементов конечное число, и их можно записать в виде последовательности V'_{i_1}, $V'_{i_2}, \ldots, V'_{i_k}$, где индекс i_ν $(\nu = 1, 2, \ldots, k)$ — номер класса, которому принадлежит V'_{i_ν}. Из (27) следует, что при $\varepsilon < 1/20$ элементы покрытия v', пересекающиеся с $\bigcap\limits_{\nu=1}^{k} V'_{i_\nu}$, также принадлежат разным классам. Пусть $V'_{j_1}, V'_{j_2}, \ldots, V'_{j_l}$ — те элементы покрытия v', которые пересекаются с $\bigcap\limits_{\nu=1}^{k} V'_{i_\nu}$, но не содержат C (j_ν — номер класса, которому принадлежит V'_{j_ν}, $\nu = 1, 2, \ldots, l$). Тогда

$$C = \left(\bigcap_{\nu=1}^{k} V'_{i_\nu}\right) \cap \left(\bigcap_{\nu=1}^{l} (H_p \setminus V'_{j_\nu})\right).$$

Таким образом, мы сопоставили каждому элементу C разбиения $\xi(v')$ символ $I(C) = (i_1, i_2, \ldots, i_k, j_1, j_2, \ldots, j_l)$, в котором все индексы различны и принимают значения от 1 до s. Так как существует лишь конечное число таких символов, осталось доказать, что если элементам C и $\tilde{C}$ отвечает один и тот же символ $I(C) = I(\tilde{C})$, то $\tilde{C}$ получается из C сдвигом на вектор, лежащий в H_p.

Пусть $I(C) = I(\tilde{C}) = (i_1, i_2, \ldots, i_k, j_1, j_2, \ldots, j_l)$, т. е.

$$C = \left(\bigcap_{\nu=1}^{k} V'_{i_\nu}\right) \cap \left(\bigcap_{\nu=1}^{l} (H_p \setminus V'_{j_\nu})\right),$$

$$\tilde{C} = \left(\bigcap_{\nu=1}^{k} \tilde{V}'_{i_\nu}\right) \cap \left(\bigcap_{\nu=1}^{l} (H_p \setminus \tilde{V}'_{j_\nu})\right),$$

где $\tilde{V}'_{i_\nu} = V'_{i_\nu} + z'_{i_\nu}$, $\tilde{V}'_{j_\nu} = V'_{j_\nu} + z'_{j_\nu}$; z'_{i_ν}, $z'_{j_\nu} \in H_p$. Покажем, что все z'_{i_ν} и z'_{j_ν} совпадают с z'_{i_1}. Пусть y'_{i_1}, y'_{j_ν} — любые точки

из V'_{i_1}, V'_{j_ν} соответственно и $\tilde{y}'_{i_1}=y'_{i_1}+z'_{i_1}$, $\tilde{y}'_{j_\nu}=y'_{j_\nu}+z'_{j_\nu}$. Из определения множеств V'_{i_ν}, V'_{j_ν} следует, что тогда $\rho\left(y'_{i_1}, y'_{j_\nu}\right)<8\varepsilon$, $\rho\left(\tilde{y}_{i_1}, \tilde{y}'_{j_\nu}\right)<8\varepsilon$, а потому $\left|z'_{i_1}-z'_{j_\nu}\right|<16\varepsilon$.

Если $z'_{i_1}\neq z'_{j_\nu}$, то при $\varepsilon<1/32$ это противоречит последнему из неравенств (27). Аналогично доказывается, что $z'_{i_\nu}=z'_{i_1}$, $\nu=1, 2, \ldots, k$. Итак, $\tilde{C}=C+z'_{i_1}$.

Следовательно, при $\varepsilon<1/32$ разбиение $\xi(v')$ обладает нужным нам свойством.

Из доказанного свойства вытекает существование такого числа $\varepsilon_1>0$, что каждый элемент C разбиения $\xi(v')$, имеющий положительную меру μ_p, содержит некоторый шар $O_{\varepsilon_1}(C)$ радиуса ε_1, лежащий в H_p. При этом шар $O_{\varepsilon_1}(C)$ для каждого C можно выбрать так, что если $\tilde{C}$ — элемент разбиения $\xi(v')$, получающийся из C сдвигом на вектор $z'\in H_p$, то $O_{\varepsilon_1}(\tilde{C})=O_{\varepsilon_1}(C)+z'$. Тогда найдутся такие числа $\varepsilon_2>0$ и $\delta<1$, что ε_2-окрестность $D_{\varepsilon_2}(C)$ границы ∂C не пересекается с $O_{\varepsilon_1}(C)$ и

$$\frac{\mu_p\left(D_{\varepsilon_2}(C)\cap C\right)}{\mu_p(C)}<\delta. \tag{28}$$

Пусть $W_i(k)$ — объединение элементов разбиения $\xi(v')$, пересекающихся с 2ε-окрестностью множества $\partial A_T^{km}V'_i(k)$. В силу неравенства (25) $\partial A_T^{km}V'_i(\infty)\subset W_i(k)$. Из определения $W_i(k)$ и оценки для $r(k, 1)$ вытекает, что

$$\sup_{y'\in A_T^{-m}W_i(k+1)}\rho\left(y', \partial A_T^{km}V'_i(k)\right)\leqslant\theta_p\sup_{y'\in W_i(k+1)}\rho\left(y', \partial A_T^{(k+1)m}V'_i(k)\right)+$$
$$+\sup_{y'\in\partial A_T^{km}V'_i(k+1)}\rho\left(y', \partial A_T^{km}V'_i(k)\right)\leqslant 6\theta_p\varepsilon+4\theta_p\varepsilon=10\theta_p\varepsilon \tag{29}$$

и, следовательно, $A_T^{-m}W_i(k+1)\subset W_i(k)$, если $\theta_p<1/7$.

Оценим отношение $\dfrac{\mu_p\left(A_T^{-m}W_i(k+1)\right)}{\mu_p(W_i(k))}$. Пусть C — элемент разбиения $\xi(v')$, принадлежащий $W_i(k)$, и $\mu_p(C)>0$. Если множество $A_T^{-m}W_i(k+1)$ пересекается с C и пересечение не содержится в $D_{\varepsilon_2}(C)$, то в $A_T^{-m}W_i(k+1)$ найдется такая точка y', что $\rho(y', \partial C)>\varepsilon_2$. По доказанному $\partial A_T^{km}V'_i(k)$ содержится в объединении границ элементов разбиения $\xi(v')$. Отсюда

следует, что $\rho(y', \partial A_T^{km} V_i'(k)) > \varepsilon_2$. Еще раз увеличив, если это необходимо, число m, можно добиться выполнения условия $10\theta_\mathrm{p}\varepsilon < \varepsilon_2$. Неравенство (29) показывает, что тогда $C \cap A_T^{-m} W_i(k+1) \subset D_{\varepsilon_2}(C)$. Используя (28), получим

$$\frac{\mu_\mathrm{p}\left(A_T^{-(k+1)m} W_i(k+1)\right)}{\mu_\mathrm{p}\left(A_T^{-km} W_i(k)\right)} = \frac{\mu_\mathrm{p}\left(A_T^{-m} W_i(k+1)\right)}{\mu_\mathrm{p}\left(W_i(k)\right)} < \delta < 1,$$

а это значит, что $\lim\limits_{k \to \infty} \mu_\mathrm{p}(A_T^{-km} W_i(k)) = 0$. Тем самым равенство (23) доказано, так как $\partial V_i'(\infty) \subset A_T^{-km} W_i(k)$ при каждом k.

Теперь можно утверждать, что множества $(V_i'(\infty), V_i''(0))$, $i = 1, 2, \ldots, s$, являются параллелограммами. Взяв их образы при отображении π, мы получим некоторое покрытие $\tilde{u}$ тора параллелограммами. Каждый элемент этого покрытия содержит соответствующий элемент покрытия u. Отсюда и из неравенства (21) вытекает, что $\tilde{u}$ — правильное покрытие. В силу (22) $\tilde{u}$ обладает одним из свойств марковского покрытия, а именно

$$\Gamma_\mathrm{c}(T^m \tilde{u}) \subset \Gamma_\mathrm{c}(\tilde{u}). \tag{30}$$

Чтобы получить настоящее марковское покрытие $\tilde{\tilde{u}}$ (одновременно правильное), достаточно повторить рассуждения этого пункта, отправляясь от $\tilde{u}$ и поменяв ролями расширяющееся и сжимающееся пространства. В самом деле, построение, превращающее $\tilde{u}$ в $\tilde{\tilde{u}}$, будет затрагивать лишь сжимающуюся границу покрытия, а потому для $\tilde{\tilde{u}}$ вместе с включением $\Gamma_\mathrm{p}(T^{-m} \tilde{\tilde{u}}) \subset \Gamma_\mathrm{p}(\tilde{\tilde{u}})$ будет выполняться и (30). Правда, число m при переходе от $\tilde{u}$ к $\tilde{\tilde{u}}$ придется еще несколько увеличить. Впрочем, из анализа изложенного доказательства видно, что подходящее m зависит лишь от начального покрытия u и, значит, может быть выбрано заранее.

§ 4. МАРКОВСКИЕ РАЗБИЕНИЯ В ДВУМЕРНОМ СЛУЧАЕ

Для автоморфизма двумерного тора можно построить конечные марковские разбиения, не прибегая к описанному выше процессу последовательных приближений. Это было замечено Адлером и Вейссом (см. [1]). Мы приведем здесь соответствующее построение. Пусть $A_T = \begin{vmatrix} m & n \\ p & q \end{vmatrix}$. Рассмотрим для простоты случай $mq - np = 1$. Начало координат O есть неподвижная точка преобразования T. Условие отсутствия

собственных значений, по модулю равных 1, в двумерном случае сводится, как нетрудно проверить, к наличию двух вещественных собственных значений λ_c, $|\lambda_c| < 1$, и λ_p, $|\lambda_p| > 1$, каждое из которых иррационально. Отвечающие им собственные направления имеют иррациональные тангенсы углов наклона.

Рассмотрим более внимательно расширяющийся и сжимающийся слои L_c и L_p, проходящие через точку O. Допустим, что $\lambda_c > 0$, $\lambda_p > 0$ и что нам удалось построить такое разбиение α на параллелограммы, что $\Gamma_c(\alpha) = \Gamma_c$ есть отрезок слоя L_c, содержащий O, а $\Gamma_p(\alpha) = \Gamma_p$ есть отрезок слоя L_p, также содержащий O. Такое разбиение непременно будет марковским, поскольку $T\Gamma_c \subset \Gamma_c$, $T^{-1}\Gamma_p \subset \Gamma_p$.

Мы сейчас увидим, что существуют разбиения тора на два параллелограмма с указанным свойством. Вообще говоря, таких разбиений счетное множество.

Будем считать слои L_c и L_p ориентированными. Рассмотрим два ориентированных отрезка $\Gamma_1 \subset L_c$ и $\Gamma_2 \subset L_p$, выходящих из точки O и пересекающихся в точке O_1 [1]), причем O_1 служит конечной точкой для обоих отрезков. Проведем затем из точки O в направлении, противоположном направлению Γ_1, отрезок Γ_3 вплоть до первого пересечения с Γ_2. Пусть это пересечение произошло в точке O_2. Ясно, что на Γ_2 точка O_2 лежит между O и O_1.

Пусть $\Gamma_c = \Gamma_1 \cup \Gamma_3$. Заметим, что Γ_c — отрезок сжимающегося слоя, содержащий точку O внутри себя. Наконец продолжим Γ_2 в первоначальном направлении вплоть до первого пересечения с Γ_c (в точке O_3) и полученный отрезок обозначим Γ_p. Нетрудно убедиться непосредственно, что отрезки Γ_c и Γ_p разбивают тор на два параллелограмма, которые мы обозначим U и U'. Расширяющиеся стороны одного параллелограмма суть отрезки OO_1 и O_2O_3, а другого — отрезки OO_2 и O_1O_3. В силу сказанного выше построенное разбиение α на два параллелограмма будет марковским.

Однако разбиение α не совсем удобно, так как элементы разбиений $T^{-k}\alpha \vee \ldots \vee T^{l}\alpha$ могут оказаться несвязными и их диаметры не будут стремиться к нулю при $k, l \to \infty$. Чтобы исправить положение, введем разбиение β, элементами которого служат связные компоненты элементов разбиения $\alpha \vee T\alpha$. Эти связные компоненты будут по-прежнему парал-

[1]) В топологической теории динамических систем такие точки, следуя Пуанкаре, называют гомоклиническими.

лелограммами, сжимающиеся границы которых содержатся в сжимающихся границах параллелограммов U, U'. Разбиение β будет марковским (см. п. 1, § 3). Нетрудно показать, что при любых $k, l \geqslant 0$ элементы разбиения $\beta_k^l = T^{-k}\beta \vee \ldots \vee T^l \beta$ представляют собой связные параллелограммы, диаметры которых стремятся к нулю при $k, l \to \infty$.

З а м е ч а н и е. Адлер и Вейсс [1] установили, что два автоморфизма двумерного тора, у которых совпадают модули собственных значений, метрически изоморфны. Известно, что существуют автоморфизмы двумерного тора с таким свойством, алгебраически неизоморфные. Из одной теоремы Арова [3] вытекает, что этот метрический изоморфизм можно установить только при помощи разрывного (но, конечно, измеримого) отображения тора на себя. Отметим, что не выяснено, к какому классу Бэра принадлежит это отображение. Перенесение результатов Адлера и Вейсса на многомерный случай является, по-видимому, трудной задачей.

ЛИТЕРАТУРА

1. Адлер, Вейсс (Adler R., Weiss B), Entropy, a complete metric invariant for automorphisms of the torus, *Proc. Nat. Acad. Sci. USA*, **57**, №6 (1967), 1573-1576.
2. А н о с о в Д. В., Геодезические потоки на замкнутых римановых многообразиях отрицательной кривизны, *Труды МИАН*, **90** (1967), 3-209.
3. А р о в Д. З., О топологическом подобии автоморфизмов и сдвигов коммутативных компактных групп, *УМН*, **18**, 5 (1963), 133-138.
4 Б у р б а к и Н., Общая топология, Физматгиз, М., 1958.
5. С и н а й Я. Г., Марковские разбиения и $У$-диффеоморфизмы, *Функц. анализ и его приложения*, **2**, 1 (1968), 64-89.
6. Ф е л л е р В., Введение в теорию вероятностей и ее приложения, „Мир“, М., 1964.

Алфавитный указатель

Оглавление

П. Биллингслей

Эргодическая теория и информация

Редактор *Л. Б. Штейнпресс*
Художник *Н. Д. Смеляков*
Художественный редактор *В. И. Шаповалов*
Технический редактор *Н. В. Соколова*

Сдано в производство 18/X 1968 г.
Подписано к печати 6/V 1969 г.
Бумага № 3 $60 \times 90^1/_{16}$=7,5 бум. л.
15 усл. печ. л.
Уч.-изд. л. 12,33. Изд. № 1/4127
Цена 85 коп. Зак. 1491

ИЗДАТЕЛЬСТВО «МИР»
Москва, 1-й Рижский пер., 2

Ленинградская типография № 2
имени Евгении Соколовой
Главполиграфпрома Комитета по печати
при Совете Министров СССР.
Измайловский проспект, 29

ИЗДАТЕЛЬСТВО „МИР“

выпускает в 1969 г.

В СЕРИИ „СОВРЕМЕННАЯ МАТЕМАТИКА“

(ПОПУЛЯРНАЯ СЕРИЯ) КНИГУ

Мостеллер Ф., Рурке Р., Томас Дж., ВЕРОЯТНОСТЬ, *перевод с английского*, издательство „Мир“, 20 изд. л.

Эта книга составлена видными американскими математиками и педагогами на основе выпущенного ранее учебника по теории вероятностей для средней школы; она предполагает у читателей весьма скромную подготовку и будет интересна и полезна всем любителям математики, начиная с увлекающихся математикой школьников старших классов. Отличительную ее черту составляет изобилие подробно разобранных примеров, содержание которых заимствовано из обыденной жизни, и задач для самостоятельного решения. Большое внимание уделяют авторы вопросам практических приложений теории вероятностей и математической статистики, которой посвящена последняя глава.

В настоящее время значение теории вероятностей и математической статистики настолько возросло, что знакомство с элементами этих наук можно считать необходимой компонентой математической культуры; настоящий учебник дает возможность приобрести эти знания.

85 к.

ИЗДАТЕЛЬСТВО «МИР»